$$\mathscr{A}+\mathscr{B}=\mathscr{B}+\mathscr{A}$$
$$(\mathscr{A}+\mathscr{B})\mathscr{C}=\mathscr{A}+(\mathscr{B}+\mathscr{C})$$
$$\mathscr{A}+\mathscr{O}=\mathscr{A}$$
$$\alpha(\mathscr{A}+\mathscr{B})=\alpha\mathscr{A}+\alpha\mathscr{B}$$
$$(\alpha+\beta)\mathscr{A}=\alpha\mathscr{A}+\beta\mathscr{A}$$
$$\mathscr{A}(\mathscr{B}+\mathscr{C})=\mathscr{A}\mathscr{B}+\mathscr{A}\mathscr{C}$$
$$(\mathscr{A}+\mathscr{B})\mathscr{C}=\mathscr{A}\mathscr{C}+\mathscr{B}\mathscr{C}$$

Fundamentos de Algebra Lineal

A.I. Máltsev

EDITORIAL MIR MOSCÚ

ИЗДАТЕЛЬСТВО
МИР

ОСНОВЫ ЛИНЕЙНОЙ АЛГЕБРЫ

А. И. Мальцев

ИЗДАТЕЛЬСТВО
«НАУКА»

A. I. Máltsev

Fundamentos de Algebra lineal

Traducido del ruso por
Carlos Vega,
candidato a Doctor en ciencias físico-matemáticas

Tercera edición

EDITORIAL
MIR
MOSCÚ

A NUESTROS LECTORES:

"Mir" edita libros soviéticos traducidos al español, inglés, francés, árabe y otros idiomas extranjeros. Entre ellos figuran las mejores obras de las distintas ramas de la ciencia y la técnica: manuales para los centros de enseñanza superior y escuelas tecnológicas; literatura sobre ciencias naturales y médicas. También se incluyen monografías, libros de divulgación científica y ciencia-ficción.
Dirijan sus opiniones a la Editorial "Mir", 1 Rizhski per., 2, 129 820, Moscú, I- I-110, GSP, URSS.

Primera edición, 1972

Segunda edición, 1976

на испанском языке

○ Traducción al español. Editorial Mir. 1978

Impreso en la URSS

INDICE

Introducción 11

Capítulo I. Matrices y determinantes

§ 1. Operaciones con matrices 13

1.1. Matrices. Campo principal (13). 1.2. Multiplicación de matrices (15). 1.3. Transposición de matrices (21). 1.4. Matrices celulares (24). 1.5. Cuaternios (28).

Ejemplos y problemas (31)

§ 2. Determinantes 33

2.1. Definición (33). 2.2. Propiedades principales de los determinantes (39). 2.3. Determinante de un producto. Matrices inversas (49). 2.4. Sistemas cramerianos de ecuaciones lineales (53).

Complementos y ejercicios (57)

§ 3. Polinomios característico y mínimo 58

3.1. Semejanza de matrices (58). 3.2. Polinomio característico (60). 3.3. Polinomio mínimo (63).

Ejemplos y problemas (67)

Capítulo II. Espacios lineales

§ 4. Dimensión 68

4.1. Módulos y espacios vectoriales (68). 4.2. Dependencia lineal (73). 4.3. Isomorfismo (81).

Ejemplos y problemas (84)

§ 5. Coordenadas 85

5.1. Coordenadas de un vector (85). 5.2. Rangos de matrices (89). 5.3. Sistemas generales de ecuaciones lineales (96).

Complementos y ejercicios (101)

§ 6. Subespacios lineales 102

6.1. Intersección y suma de subespacios (102). 6.2. Sumas directas (107). 6.3. Sistemas de ecuaciones lineales homogéneas (109).

Ejemplos y problemas (113)

Capítulo III. Aplicaciones lineales

§ 7. Aplicaciones de conjuntos arbitrarios . . . 114

7.1. Producto de aplicaciones (114). 7.2. Las aplicaciones idéntica e inversa (116). 7.3. Aplicaciones biyectivas (117). 7.4. Sustituciones (118).

Ejemplos y problemas (121)

§ 8. Aplicaciones lineales y sus matrices 121
8.1. Propiedades elementales (121). 8.2. Matriz de una aplicación lineal (124). 8.3. Transformación de coordenadas (125).
Ejemplos y problemas (127)

§ 9. Operaciones con aplicaciones lineales 127
9.1. Multiplicación de aplicaciones lineales (127). 9.2. Multiplicación por número y adición (129). 9.3. Polinomios en aplicaciones lineales (131).
Ejemplos y problemas (132)

§ 10. Rango y defecto de una aplicación lineal 132
10.1. Núcleo y dominio de valores (132). 10.2. Aplicaciones singulares y regulares (135). 10.3. Rango de la matriz de una aplicación (137).
Ejemplos y problemas (138)

§ 11. Subespacios invariantes 139
11.1. Aplicación inducida (139). 11.2. Suma directa de subespacios invariantes (141). 11.3. Polinomio característico de una aplicación (143). 11.4. Vectores propios y valores propios (144).
Ejemplos y problemas (147)

§ 12. Aplicaciones de matrices de forma normal 147
12.1. Forma diagonal (147). 12.2. Células de Jordan (149). 12.3. Subespacios radicales (150).
Ejemplos y problemas (152)

Capítulo IV. Matrices polinomiales

§ 13. Factores invariantes 153
13.1. Equivalencia (153). 13.2. Forma diagonal (155). 13.3. Máximos comunes divisores de menores (158). 13.4. Condiciones de equivalencia (162).
Ejemplos y problemas (165)

§ 14. Divisores elementales 166
14.1. Relación con los factores invariantes (166). 14.2. Divisores elementales de una matriz descompuesta (167).
Ejemplos y problemas (169)

§ 15. Formas normales de la matriz de una aplicación lineal 170
15.1. División de λ-matrices (170). 15.2. Equivalencia escalar. (171). 15.3. Criterio de semejanza de matrices

(172). 15.4. Forma normal de Jordan (174). 15.5. Forma normal natural (177). 15.6. Otras formas normales (178).

Ejemplos y problemas (181)

§ 16. Funciones de matrices 183

16.1. Polinomio en una matriz de Jordan (183). 16.2. Funciones escalares (184). 16.3. Representación de los valores de funciones por polinomios (187). 16.4. Divisores elementales de funciones (189). 16.5. Series de potencias (192). 16.6. Matrices conmutables con una matriz dada (193). 16.7. Matrices que conmutan con matrices conmutables (197).

Ejemplos y problemas (199)

Capítulo V. Espacios unitarios y euclídeos

§ 17. Espacios unitarios 200

17.1. Axiomática y ejemplos (200). 17.2. Longitud de un vector (204). 17.3. Sistemas ortonormales (206). 17.4. Isomorfismo (211). 17.5. Sumas ortogonales. Proyecciones (212).

Ejemplos y problemas (214)

§ 18. Aplicaciones conjugadas 215

18.1. Funciones lineales (215). 18.2. Aplicaciones conjugadas (218). 18.3. Aplicaciones normales (220).

Ejemplos y problemas (225)

§ 19. Aplicaciones unitarias y simétricas 225

19.1. Aplicaciones unitarias (225). 19.2. Equivalencia unitaria (228). 19.3. Forma normal de la matriz de una aplicación unitaria (230). 19.4. Aplicaciones simétricas (231). 19.5. Aplicaciones antisimétricas (233). 19.6. Aplicaciones simétricas no negativas (235).

Ejemplos y problemas (239)

§ 20. Descomposición de aplicaciones generales . . 240

20.1. Descomposición en partes simétrica y antisimétrica (240). 20.2. Descomposición polar (241). 20.3. Aplicación de Cayley (245). 20.4. Descomposición espectral (247)

Ejemplos y problemas (252)

Capítulo VI. Formas bilineales y cuadráticas

§ 21. Formas bilineales 254

21.1. Transformación de formas (254). 21.2. Equivalencia de formas bilineales (256). 21.3. Congruencia de formas bilineales simétricas (259).

Ejemplos y problemas (261)

§ 22. Formas cuadráticas 262

22.1. Congruencia (262). 22.2. Algoritmo de Lagrange (264). 22.3. Ley de inercia de formas cuadráticas (268). 22.4. Formas de signo constante (269).

Ejemplos y problemas (271)

§ 23. Pares de formas 271

23.1. Equivalencia de pares de formas (271). 23.2. Congruencia de pares de formas (273). 23.3. Congruencia de formas bilineales no simétricas (276).

Ejemplos y problemas (278)

§ 24. Funciones bilineales 278

24.1. Definiciones principales (278). 24.2. Espacios de métrica bilineal (282). 24.3. Funciones bilineales en espacios bilineales métricos (286).

Ejemplos y problemas (291)

Capítulo VII. Aplicaciones lineales de espacios bilineales métricos

§ 25. Tipos principales de aplicaciones lineales . . 293

25.1. Automorfismos (293). 25.2. Aplicaciones simétricas y antisimétricas (298).

Ejemplos y problemas (300)

§ 26. Espacios euclídeos complejos 300

26.1. Aplicaciones simétricas (301). 26.2. Aplicaciones antisimétricas (303). 26.3. Aplicaciones ortogonales complejas (306).

Ejemplos y problemas (308)

§ 27. Espacios simpliciales 309

27.1. Aplicaciones simétricas (309). 27.2. Aplicaciones antisimétricas (311). 27.3. Aplicaciones simpliciales (313).

Ejemplos y problemas (314)

§ 28. Espacios seudounitarios 315
28.1. Aplicaciones simétricas (315). 28.2. Aplicaciones seudounitarias (323).
Ejemplos y problemas (324)

Capítulo VIII. Espacios afines

§ 29. Espacios afines generales 326
29.1. Axiomática (326). 29.2. Variedades lineales (335). 29.3. Planos paralelos (344). 29.4. Funcionales lineales (346).
Complementos y ejemplos (351)

§ 30. Coordenadas afines 352
30.1. Coordenadas de un punto (352). 30.2. Ecuaciones de planos (355). 30.3. Ecuaciones de hiperplanos y de rectas (363). 30.4. Transformación de coordenadas afines (368).
Ejemplos y problemas (373)

§ 31. Cuerpos convexos 373
31.1. Rayos (374). 31.2. Semiespacios (377). 31.3. Conjuntos convexos (380).
Complementos y ejemplos (385)

§ 32. Espacios euclídeos puntuales 386
32.1. Longitud de una quebrada (386). 32.2. Angulo entre rectas (388). 32.3. Proyecciones ortogonales (391). 32.4. Angulo entre un plano y una recta (396).
Ejemplos y problemas (398)

Indice alfabético 399

INTRODUCCION

El Algebra lineal es una rama de las Matemáticas tan antigua como la propia Matemática. El problema de la solución de la ecuación lineal $ax+b=0$ puede ser considerado como el problema primario del Algebra lineal. Aunque este problema no representa dificultad alguna, el método de su solución, así como las propiedades de la función lineal correspondiente $y=ax+b$, son los modelos de partida para las ideas y los métodos de toda el álgebra lineal. Por ejemplo, la teoría de la solución de un sistema de ecuaciones con varias incógnitas se basa en la idea de la sustitución del sistema dado por una cadena de ecuaciones del tipo indicado y de la forma más sencilla.

La importancia de los sistemas de ecuaciones lineales aumentó particularmente con la creación de la Geometría analítica, que permitió reducir todos los problemas principales sobre la posición de planos y rectas en el espacio al estudio de sistemas de ecuaciones lineales. Ya en el siglo XVIII la búsqueda de fórmulas generales para la solución de un sistema de n ecuaciones con n incógnitas llevó a Leibniz y a Cramer al concepto de determinante. En el siglo XIX, además del Algebra y de la Geometría analítica, los determinantes penetraron también en el Análisis con los trabajos de Ostrogradski, Jacobi (determinantes funcionales), Wronski y otros. Paralelamente en la Geometría analítica, en la teoría de los números y especialmente en la mecánica teórica adquiría cada vez mayor importancia el problema de transformación de formas cuadráticas mediante sustituciones lineales de las variables. Este problema resultó ser también uno de los centrales en el desarrollo de las ideas geométricas de Lobachevski y de Riemann, que llevó a la creación de la teoría de espacios lineales multidimensionales (Grassmann). A mediados del siglo pasado y en relación con el estudio de álgebras no conmutativas (Hamilton) apareció en los trabajos de Cayley y de Sylvester el cálculo de matrices, que en el desarrollo posterior del Algebra lineal pasó a ocupar uno de los puestos principales. A finales del siglo XIX quedaron creados los capítulos principales del cálculo de matrices: forma normal de una matriz de una transformación lineal (Jordan), divisores elementales (Weierstrass), pares de formas cuadráticas (Weierstrass, Kronecker), formas hermitianas (Hermite). El desarrollo de la geometría diferencial de espacios multidimensionales y de la teoría de transformaciones de formas algebraicas de órdenes superiores llevó, a finales del siglo XIX, a la creación del cálculo tensorial.

En el siglo actual los métodos del Algebra lineal han encontrado amplia aplicación y han sido desarrollados en la teoría de anillos y módulos, en la teoría de representaciones de grupos, así como en

la teoría de espacios topológicos vectoriales y otros capítulos del Análisis funcional. Ya en las dos últimas décadas la teoría de desigualdades lineales y la teoría de espacios afines multidimensionales, estrechamente ligada a la primera, han ocupado uno de los lugares centrales en una rama tan conocida de la matemática aplicada como es la teoría de operaciones. Gracias a ello los elementos de la teoría de espacios afines multidimensionales constituyen ahora un momento indispensable en la formación matemática de ingenieros y economistas.

En el Algebra lineal se estudian objetos de tres géneros: matrices, espacios y formas algebraicas. Las teorías de estos objetos están estrechamente vinculadas. La mayoría de los problemas del Algebra lineal admite un enunciado natural en términos de cada una de las tres teorías señaladas. El enunciado matricial es generalmente el más cómodo para los fines de cálculo. Por otra parte, en la geometría y en la mecánica la mayoría de los problemas del Algebra lineal aparece como problemas de estudio de formas algebraicas. Sin embargo, la comprensión más clara de las relaciones internas de diferentes problemas del Algebra lineal se alcanza solamente al considerar los espacios lineales correspondientes que son por ello el objeto principal de estudio en el Algebra lineal.

Desde el punto de vista de la teoría de formas el contenido del Algebra lineal se descompone de modo natural en la teoría de formas lineales, cuadráticas y de órdenes superiores. El álgebra lineal propiamente dicha se relaciona, en general, solamente con la teoría de formas lineales y cuadráticas, así como con los elementos iniciales de la teoría de formas polilineales y del álgebra tensorial.

Capítulo I

Matrices y determinantes

§ 1. Operaciones con matrices

1.1. Matrices. Campo principal. Los objetos principales de estudio en lo sucesivo serán las matrices, los espacios lineales y los polinomios de varias variables, llamados también formas algebraicas. En la definición de cada uno de estos objetos participa un conjunto K de números o elementos de otra índole que debe ser previamente elegido. La elección concreta de K depende de los problemas que se resuelven y de la disciplina científica. Por ejemplo, desde el punto de vista algebraico los resultados obtienen la forma más completa, si K es el conjunto de todos los números complejos. Por el contrario, en la geometría y en la mecánica es preciso considerar generalmente los números reales, mientras que en la teoría de los números resulta natural aceptar que K es el conjunto de los números racionales o, incluso, solamente el conjunto de los números racionales enteros. Para que los resultados puedan ser aplicados a un número de problemas lo más amplio posible conviene, por lo tanto, no fijar de antemano qué conjunto concreto se comprende por K. En algunas secciones será suficiente suponer que K es un anillo asociativo. En varios capítulos aceptaremos que K es un cuerpo conmutativo arbitrario o un cuerpo conmutativo arbitrario ordenado, mientras que varios teoremas importantes serán demostrados solamente en la suposición de que K es el conjunto de todos los números reales o el conjunto de todos los números complejos. Para las aplicaciones geométricas y físicas son de mayor importancia precisamente los casos en que K es el cuerpo de los números reales o el cuerpo de los números complejos.

Los elementos del conjunto K se llamarán *números* incluso cuando K sea un anillo arbitrario. Los representaremos por letras griegas minúsculas α, β, ..., τ.

Un sistema arbitrario de elementos del conjunto K dispuestos en forma de una tabla rectangular de m filas y de n columnas se denomina (m, n)-matriz o simplemente *matriz* sobre K. Para

representar una matriz, los símbolos que designan sus elementos suelen escribirse en el orden adecuado y la tabla obtenida se incluye entre paréntesis, corchetes o barras verticales dobles. Por consiguiente, la forma general de una (m, n)-matriz será

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2n} \\ \dots & \dots & \dots & \dots \\ \alpha_{m1} & \alpha_{m2} & \dots & \alpha_{mn} \end{bmatrix}, \quad \begin{pmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2n} \\ \dots & \dots & \dots & \dots \\ \alpha_{m1} & \alpha_{m2} & \dots & \alpha_{mn} \end{pmatrix}, \quad \begin{Vmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2n} \\ \dots & \dots & \dots & \dots \\ \alpha_{m1} & \alpha_{m2} & \dots & \alpha_{mn} \end{Vmatrix},$$

donde α_{ij} representan elementos de K. En vez de esta notación detallada con frecuencia se emplea la notación abreviada: $\|\alpha_{ij}\|$ o $\|\alpha_{ij}\|_{m,n}$.

Si el número de filas coincide con el número de columnas, la matriz se llama *cuadrada* y el número de sus filas, igual al de sus columnas, se llama *orden* de la matriz cuadrada. En particular, una matriz cuadrada de orden 1 es simplemente un elemento de K.

Una matriz compuesta de una sola fila se llama simplemente *fila* y el número de sus elementos se denomina *longitud* de fila. En lo sucesivo las matrices serán representadas por letras mayúsculas latinas.

Dos matrices se llaman *iguales*, si son iguales los números de sus filas y de sus columnas respectivamente y si coinciden los números que ocupan posiciones correspondientes en estas matrices. Por consiguiente, una igualdad entre dos (m, n)-matrices equivale a mn igualdades entre sus elementos.

Las operaciones matriciales principales son: la multiplicación de un número por una matriz o de una matriz por un número, la adición y la multiplicación de dos matrices. Por definición, para multiplicar el número α por la matriz A o la matriz A por el número α hay que multiplicar por α todos los elementos de la matriz A. Por ejemplo,

$$\alpha \cdot \begin{bmatrix} \lambda & \mu \\ \xi & \eta \end{bmatrix} = \begin{bmatrix} \alpha\lambda & \alpha\mu \\ \alpha\xi & \alpha\eta \end{bmatrix}, \quad \begin{bmatrix} \lambda & \mu \\ \xi & \eta \end{bmatrix} \cdot \alpha = \begin{bmatrix} \lambda\alpha & \mu\alpha \\ \xi\alpha & \eta\alpha \end{bmatrix}.$$

Si las matrices se consideran sobre un anillo conmutativo K, es válida la igualdad $\alpha A = A\alpha$ para cualquier matriz A y para cualquier $\alpha \in K$. En el caso de un anillo no conmutativo K puede resultar que $\alpha A \neq A\alpha$. Siendo K el anillo de todos los números enteros, tenemos, por ejemplo,

$$5 \cdot \begin{bmatrix} 2 & 3 \\ 7 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 7 & -1 \end{bmatrix} \cdot 5 = \begin{bmatrix} 10 & 15 \\ 35 & -5 \end{bmatrix}.$$

La matriz, en la cual todos los elementos son iguales a cero, se llama *matriz nula* y se designa O. Si se quiere indicar de manera explícita el número de filas y de columnas de la matriz nula, se escribe O_{mn}.

Está claro que para toda matriz A sobre K y para cualesquiera $\alpha, \beta \in K$ tienen lugar las relaciones:

1. $1 \cdot A = A \cdot 1 = A$.
2. $0 \cdot A = A \cdot 0 = 0$; $\alpha \cdot 0 = 0 \cdot \alpha = 0$.
3. $\alpha(\beta A) = (\alpha\beta) A$; $(A\alpha)\beta = A(\alpha\beta)$.

Se llama suma de dos matrices A y B de igual cantidad de filas y de columnas respectivamente una matriz con el mismo número de filas y de columnas, cuyos elementos son iguales a las sumas de los elementos correspondientes de las matrices A y B. Por ejemplo,

$$\begin{bmatrix} 2 & 1 & 3 \\ 1 & -3 & 5 \end{bmatrix} + \begin{bmatrix} 0 & 2 & -2 \\ 1 & 2 & -5 \end{bmatrix} = \begin{bmatrix} 2 & 3 & 1 \\ 2 & -1 & 0 \end{bmatrix}.$$

De esta definición se desprenden inmediatamente las relaciones:

4. $A + (B + C) = (A + B) + C$;
5. $A + B = B + A$;
6. $(\alpha + \beta) A = \alpha A + \beta A$; $A(\alpha + \beta) = A\alpha + A\beta$;
7. $\alpha(A + B) = \alpha A + \alpha B$; $(A + B)\alpha = A\alpha + B\alpha$;
8. $A + 0 = 0 + A = A$;

la demostración queda a cargo del lector. En particular, empleando las propiedades 1 y 6, obtenemos

$$A + A = 2A, \quad A + A + A = 3A, \quad \ldots$$

Introduciendo la notación $(-1) A = -A$, tendremos también

$$A + (-A) = 0, \quad (-\alpha) A = -\alpha A, \quad -(A + B) = -A - B,$$
$$-(-A) = A.$$

Para abreviar, en lugar de $A + (-B)$ suele escribirse $A - B$.

1.2. Multiplicación de matrices. A diferencia de las operaciones de adición y de multiplicación por un número, la operación de multiplicación de una matriz por otra se define de forma más compleja. A saber, sean dadas dos matrices A y B, tales que el número de columnas de la primera coincide con el número de filas de la segunda. Si

$$A = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \ldots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \ldots & \alpha_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ \alpha_{m1} & \alpha_{m2} & \ldots & \alpha_{mn} \end{bmatrix} \quad \text{y} \quad B = \begin{bmatrix} \beta_{11} & \beta_{12} & \ldots & \beta_{1p} \\ \beta_{21} & \beta_{22} & \ldots & \beta_{2p} \\ \ldots & \ldots & \ldots & \ldots \\ \beta_{n1} & \beta_{n2} & \ldots & \beta_{np} \end{bmatrix},$$

la matriz

$$C = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \ldots & \gamma_{1p} \\ \gamma_{21} & \gamma_{22} & \ldots & \gamma_{2p} \\ \ldots & \ldots & \ldots & \ldots \\ \gamma_{m1} & \gamma_{m2} & \ldots & \gamma_{mp} \end{bmatrix},$$

donde

$$\gamma_{ij} = \alpha_{i1}\beta_{1j} + \alpha_{i2}\beta_{2j} + \ldots + \alpha_{in}\beta_{nj} \quad (i = 1, \ldots, m;\ j = 1, \ldots, p),$$

se denomina *producto de A por B* y se designa AB. Por ejemplo,

$$\begin{bmatrix} \alpha & \beta \\ \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{bmatrix} \begin{bmatrix} \gamma & \delta & \varepsilon \\ \lambda & \mu & \nu \end{bmatrix} = \begin{bmatrix} \alpha\gamma + \beta\lambda & \alpha\delta + \beta\mu & \alpha\varepsilon + \beta\nu \\ \alpha_1\gamma + \beta_1\lambda & \alpha_1\delta + \beta_1\mu & \alpha_1\varepsilon + \beta_1\nu \\ \alpha_2\gamma + \beta_2\lambda & \alpha_2\delta + \beta_2\mu & \alpha_2\varepsilon + \beta_2\nu \end{bmatrix}.$$

La regla de multiplicación de matrices se enuncia, a veces, de la siguiente forma: *para obtener el elemento, que se encuentra en la i-ésima fila y j-ésima columna del producto de dos matrices, hay que multiplicar los elementos de la i-ésima fila de la primera matriz por los elementos correspondientes de la j-ésima columna de la segunda matriz y sumar los productos obtenidos.*

El producto de dos matrices, hablando en términos generales, depende del orden de los factores incluso en el caso en que el anillo K es conmutativo. Por ejemplo,

$$\begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 4 \\ 5 & 8 \end{bmatrix},$$

$$\begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 7 & 6 \\ 4 & 4 \end{bmatrix}.$$

Si se consideran matrices no cuadradas, puede ocurrir incluso que el producto de dos matrices tomadas en un orden tenga sentido y tomadas en el orden contrario, no lo tenga.

Demostremos ahora las propiedades principales de la multiplicación de matrices.

9. $\alpha(AB) = (\alpha A)B;\quad A(\alpha B) = (A\alpha)B;\quad (AB)\alpha = A(B\alpha).$

Sean $A = \|\alpha_{ij}\|_{mn}$ y $B = \|\beta_{jk}\|_{np}$. Para el elemento que se encuentra en la i-ésima fila y k-ésima columna de la matriz $\alpha(AB)$ $(i = 1, \ldots, m;\ k = 1, \ldots, p)$, obtenemos, empleando la regla de multiplicación de matrices, la expresión siguiente:

$$\alpha(\alpha_{i1}\beta_{1k} + \ldots + \alpha_{in}\beta_{nk}).$$

Análogamente, para el elemento que se encuentra en la misma i-ésima fila y en la misma k-ésima columna de la matriz $(\alpha A)B$, obtenemos la expresión

$$(\alpha\alpha_{i1})\beta_{1k} + \ldots + (\alpha\alpha_{in})\beta_{nk}.$$

Como ambas expresiones coinciden, queda demostrada la primera de las igualdades 9. Con cálculos semejantes se demuestran las otras dos igualdades de 9, así como las propiedades:

10. $(A + B)C = AC + BC.$
11. $C(A + B) = CA + CB.$

De las propiedades 10 y 11 se desprende directamente la siguiente regla general: *para multiplicar una suma de matrices por otra hay que multiplicar cada matriz de la primera suma por cada matriz de la segunda suma y sumar los productos obtenidos.*

Hemos visto que para el producto de matrices no se cumple la ley conmutativa: AB puede ser distinto de BA. Sin embargo, la segunda ley aritmética—la ley asociativa de la multiplicación—se cumple para la multiplicación de matrices[1].

12. $A(BC)=(AB)C$.

Para la demostración tomemos

$$AB=M \quad \text{y} \quad BC=N$$

y representemos mediante μ_{ik} y ν_{jl} los elementos de las matrices M y N. Según la regla de multiplicación de matrices tenemos

$$\mu_{ik}=\alpha_{i1}\beta_{1k}+\alpha_{i2}\beta_{2k}+\ldots+\alpha_{in}\beta_{nk},$$
$$\nu_{jl}=\beta_{j1}\gamma_{1l}+\beta_{j2}\gamma_{2l}+\ldots+\beta_{jp}\gamma_{pl},$$

donde α_{ij}, β_{jk} y γ_{kl} son los elementos de las matrices A, B y C. Efectuando la multiplicación de M por C, obtendremos en la i-ésima fila y l-ésima columna de la matriz $(AB)C$ la suma

$$\mu_{i1}\gamma_{1l}+\mu_{i2}\gamma_{2l}+\ldots+\mu_{ip}\gamma_{pl}=\sum_k\sum_j\alpha_{ij}\beta_{jk}\gamma_{kl}.$$

Análogamente, efectuando la multiplicación de A por N, obtendremos en la i-ésima fila y l-ésima columna del producto $A(BC)$ la suma

$$\alpha_{i1}\nu_{1l}+\alpha_{i2}\nu_{2l}+\ldots+\alpha_{in}\nu_{nl}=\sum_j\sum_k\alpha_{ij}\beta_{jk}\gamma_{kl}.$$

Puesto que estas dos sumas difieren solamente en el orden de los sumandos, la fórmula 12 queda demostrada.

De la fórmula 12 se deduce que el producto de varias matrices dispuestas en un orden determinado no depende de cómo se coloquen los paréntesis. Por esto podemos hablar no sólo sobre el producto de dos matrices, sino también sobre el producto de un número mayor de matrices. Por ejemplo, podemos hablar simplemente del producto $ABCD$ de cuatro matrices, ya que las cinco formas diferentes de calcular este producto

$$((AB)C)D, \quad (A(BC))D, \quad A((BC)D), \quad A(B(CD)), \quad (AB)(CD)$$

llevan al mismo resultado. En efecto, cada producto siguiente se obtiene del anterior aplicando directamente la ley asociativa 12.

[1] Puesto que no se pueden sumar y multiplicar matrices arbitrarias, sino solamente aquellas en las que el número de filas y de columnas está sujeto a condiciones determinadas, las igualdades 10, 11 y 12 deben comprenderse de manera que si las operaciones indicadas en uno de los miembros son posibles, las operaciones indicadas en el otro miembro también son posibles y los resultados obtenidos en ambos miembros coinciden.

Ya hemos señalado que cualesquiera dos matrices no pueden ser sumadas o multiplicadas, ya que para poder realizar estas operaciones es preciso que se cumplan determinadas relaciones entre los números de filas y de columnas. Este inconveniente desaparece si se consideran solamente matrices cuadradas de un orden fijo n. Cualesquiera dos matrices de este tipo pueden ser sumadas o multiplicadas, así como multiplicadas por cualesquiera números de K, y el resultado será otra vez una matriz cuadrada del mismo orden n. Las propiedades 1—12 indican que *el conjunto de todas las matrices cuadradas de orden dado n sobre un anillo asociativo arbitrario K forman, a su vez, un anillo asociativo respecto a las operaciones matriciales de adición y de multiplicación.*

En lo que sigue aceptaremos que el conjunto numérico principal K es un anillo asociativo con el elemento unidad 1. La matriz cuadrada, en la que todos los elementos diagonales son iguales a 1 y los restantes son iguales a cero, se denomina matriz *unidad* y se designa mediante E o E_n, donde n es su orden. Por consiguiente,

$$E=\begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

Efectuando el cálculo directo, es fácil obtener para cualquier matriz cuadrada A la igualdad

$$AE=EA=A,$$

que expresa la propiedad fundamental de la matriz E. Las matrices de tipo

$$\begin{bmatrix} \alpha & 0 & \dots & 0 \\ 0 & \beta & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & \gamma \end{bmatrix},$$

se denominan *diagonales*.

De las reglas para las operaciones se desprende directamente que la suma y el producto de matrices diagonales son también matrices diagonales:

$$\begin{bmatrix} \alpha & 0 & \dots & 0 \\ 0 & \beta & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & \gamma \end{bmatrix}+\begin{bmatrix} \alpha_1 & 0 & \dots & 0 \\ 0 & \beta_1 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & \gamma_1 \end{bmatrix}=\begin{bmatrix} \alpha+\alpha_1 & 0 & \dots & 0 \\ 0 & \beta+\beta_1 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & \gamma+\gamma_1 \end{bmatrix},$$

$$\begin{bmatrix} \alpha & 0 & \dots & 0 \\ 0 & \beta & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & & \gamma \end{bmatrix} \cdot \begin{bmatrix} \alpha_1 & 0 & \dots & 0 \\ 0 & \beta_1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \gamma_1 \end{bmatrix} = \begin{bmatrix} \alpha\alpha_1 & 0 & & 0 \\ 0 & \beta\beta_1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \gamma\gamma_1 \end{bmatrix}.$$

Consideremos ahora una matriz cuadrada cualquiera X de orden n formada por elementos del anillo K. Por definición, tomamos

$$X^0 = E, \quad X^1 = X, \quad X^2 = XX, \quad X^3 = XXX, \dots$$

Puesto que en los productos de varias matrices los paréntesis pueden ser dispuestos arbitrariamente, tenemos para cualesquiera enteros no negativos p y q y para cualquier matriz cuadrada X sobre el anillo asociativo K

$$X^p X^q = X^{p+q}, \tag{1}$$

$$(X^p)^q = X^{pq}. \tag{2}$$

Las matrices A y B se llaman *conmutables*, si

$$AB = BA. \tag{3}$$

De la relación (1) obtenemos

$$X^p X^q = X^{p+q} = X^q X^p,$$

y, por consiguiente, *todas las potencias naturales de una misma matriz son conmutables entre sí.*

Es válida incluso una afirmación más general: *si las matrices A y B son conmutables, cualesquiera potencias naturales de las mismas también son conmutables y para cualquier p natural se tiene*

$$(AB)^p = A^p B^p. \tag{4}$$

En efecto, para cualesquiera naturales p y q se tiene

$$A^p B^q = AA \dots AB \dots B.$$

Por hipótesis, en este producto pueden ser permutados cualesquiera dos factores contiguos. Pero mediante permutaciones de este tipo los factores pueden ser dispuestos en cualquier orden y, en particular, todos los factores iguales a B pueden ser llevados a las posiciones primeras. Análogamente se demuestra también la fórmula (4).

Consideremos ahora un polinomio

$$\varphi(\lambda) = \alpha_0 + \alpha_1\lambda + \dots + \alpha_n\lambda^n$$

en la letra λ, cuyos coeficientes pertenecen al anillo K. Si A es una matriz cuadrada sobre K, la expresión

$$\alpha_0 E + \alpha_1 A + \dots + \alpha_n A^n$$

se denomina *valor* del polinomio $\varphi(\lambda)$ para $\lambda = A$ o, brevemente, polinomio correspondiente en la matriz A. Suponiendo que el anillo K es conmutativo, llegamos fácilmente a la conclusión de que *el valor de una suma de polinomios en λ para $\lambda = A$ es igual a la suma de los valores de los sumandos y de que el valor de un producto de polinomios es igual al producto de los valores de los factores.*

Sean

$$\varphi(\lambda) = \alpha_0 + \alpha_1\lambda + \ldots + \alpha_n\lambda^n,$$
$$\psi(\lambda) = \beta_0 + \beta_1\lambda + \ldots + \beta_n\lambda^n.$$

Entonces,

$$f(\lambda) = \varphi(\lambda) + \psi(\lambda) = (\alpha_0 + \beta_0) + (\alpha_1 + \beta_1)\lambda + \ldots + (\alpha_n + \beta_n)\lambda^n,$$
$$g(\lambda) = \varphi(\lambda)\psi(\lambda) = \alpha_0\beta_0 + (\alpha_0\beta_1 + \alpha_1\beta_0)\lambda + \ldots + \alpha_n\beta_n\lambda^{2n}.$$

Nuestras afirmaciones consisten en que

$$f(A) = \varphi(A) + \psi(A),$$
$$g(A) = \varphi(A)\psi(A).$$

Para la demostración es suficiente escribir las expresiones para $\varphi(A)$, $\psi(A)$, $f(A)$ y $g(A)$ y siguiendo las reglas 1—12 del cálculo matricial efectuar la adición y la multiplicación de $\varphi(A)$ y $\psi(A)$.

A título de ejemplo consideremos la igualdad

$$\lambda^3 - 1 = (\lambda - 1)(\lambda + 1).$$

Tomando los valores del primero y segundo miembro para $\lambda = A$, obtenemos la igualdad matricial

$$A^3 - E = (A - E)(A + E).$$

De manera análoga de la igualdad

$$\lambda^3 + 1 = (\lambda + 1)(\lambda^2 - \lambda + 1)$$

obtenemos la relación

$$A^3 + E = (A + E)(A^2 - A + E).$$

En general, de esta forma se puede obtener de toda relación entre polinomios en λ una identidad matricial. En particular, según las reglas de operaciones con polinomios, se tiene

$$\varphi(\lambda)\psi(\lambda) = \psi(\lambda)\varphi(\lambda).$$

Tomando aquí en lugar de λ una matriz cuadrada A, obtenemos

$$\varphi(A)\psi(A) = \psi(A)\varphi(A).$$

Por consiguiente, los polinomios en una misma matriz son conmutables.

1.3. Transposición de matrices. Consideremos una matriz arbitraria

$$A = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2n} \\ \dots & \dots & \dots & \dots \\ \alpha_{m1} & \alpha_{m2} & \dots & \alpha_{mn} \end{bmatrix}.$$

La matriz

$$A' = \begin{bmatrix} \alpha_{11} & \alpha_{21} & \dots & \alpha_{m1} \\ \alpha_{12} & \alpha_{22} & \dots & \alpha_{m2} \\ \dots & \dots & \dots & \dots \\ \alpha_{1n} & \alpha_{2n} & \dots & \alpha_{mn} \end{bmatrix},$$

que se obtiene de A al cambiar las filas por las columnas, se llama *transpuesta* respecto de A. En lo sucesivo la raya siempre indicará el paso a la matriz transpuesta.

Para dos matrices arbitrarias A y B tienen lugar las siguientes reglas de transposición:

$$(\alpha A + \beta B)' = \alpha A' + \beta B',$$
$$(AB)' = B'A',$$

donde α y β son números cualesquiera. Demostremos, por ejemplo, la segunda de estas igualdades. El elemento, que se encuentra en la i-ésima fila y j-ésima columna de la matriz $(AB)'$, es igual al elemento que aparece en la j-ésima fila e i-ésima columna de la matriz AB, es decir, es igual a

$$\alpha_{j1}\beta_{1i} + \alpha_{j2}\beta_{2i} + \dots + \alpha_{jn}\beta_{ni},$$

donde α_{ij} y β_{ij} son los elementos de las matrices A y B. Pero esta expresión es la suma de los productos de los elementos de la i-ésima fila de la matriz B' por los elementos correspondientes de la j-ésima columna de la matriz A'; es decir, $(AB)' = B'A'$.

Si A es una matriz cuadrada cualquiera y

$$A' = A,$$

se dice que A *es simétrica*; en cambio, si

$$A' = -A,$$

se dice que A es *antisimétrica*. Los elementos simétricos respecto de la diagonal principal coinciden en una matriz simétrica y son opuestos en una matriz antisimétrica. En particular, todos los elementos diagonales de una matriz antisimétrica son iguales a cero.

De la regla de transposición de una suma se deduce directamente que la suma de matrices simétricas es una matriz simétrica y que la suma de matrices antisimétricas es una matriz antisimétrica. El

producto de matrices simétricas puede no ser una matriz simétrica; por ejemplo,

$$\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} \cdot \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 3 \\ 7 & 5 \end{bmatrix}.$$

Sin embargo, si dos matrices simétricas A y B son conmutables, el producto también será una matriz simétrica. En efecto, en este caso se tiene

$$(AB)' = B'A' = BA = AB.$$

De aquí se deduce que las potencias de una matriz simétrica son matrices simétricas y que los polinomios en una matriz simétrica también son matrices simétricas.

Una matriz cuadrada A sobre el anillo K se llama *invertible* (sobre K), si existe una matriz cuadrada X sobre K, que satisface las relaciones

$$AX = XA = E. \tag{1}$$

Toda matriz X que verifica las condiciones (1) se denomina matriz *inversa* de A o *inversión* de la matriz A. Para toda matriz invertible A existe solamente una inversión. En efecto, si además de la matriz X hay otra matriz Y que satisface las condiciones (1), multiplicando a la izquierda por X ambos miembros de la igualdad

$$AY = E,$$

obtenemos

$$XA \cdot Y = XE$$

o $Y = X$.

La inversión de la matriz A, si es que existe, se designa mediante A^{-1}. Es decir, por definición,

$$A \cdot A^{-1} = A^{-1} \cdot A = E. \tag{2}$$

En las condiciones (1) las matrices A y X figuran simétricamente y, por ello, si X es la inversión de A, A es la inversión de X; en otras palabras,

$$(A^{-1})^{-1} = A. \tag{3}$$

Si las matrices cuadradas A, B y C son de un mismo orden e invertibles, su producto ABC también es invertible y

$$(ABC)^{-1} = C^{-1}B^{-1}A^{-1},$$

es decir, *la inversión de un producto de matrices es igual al producto de las inversiones de los factores tomado en el orden contrario.*

Para la demostración es preciso comprobar las igualdades

$$ABC \cdot C^{-1}B^{-1}A^{-1} = C^{-1}B^{-1}A^{-1} \cdot ABC = E,$$

que son consecuencias evidentes de las relaciones (2) y de las relaciones análogas para las matrices B y C.

Para toda matriz invertible A, además de sus potencias naturales $A^0 = E$, $A^1 = A$, $A^2 = AA$, ..., se consideran también sus potencias negativas enteras, tomando por definición

$$A^{-2} = A^{-1}A^{-1}, \quad A^{-3} = A^{-1}A^{-1}A^{-1}, \quad \ldots \tag{4}$$

Las potencias fraccionarias de matrices se consideran raramente, debido a que en muchos casos las definiciones corrientes no ofrecen valores unívocos para estas potencias (véase el p. 16.2 del cap. IV).

De las relaciones (2) y (4) se deduce que para cualquier matriz invertible A y cualesquiera números enteros (no necesariamente positivos) p y q tienen lugar las reglas comunes de operaciones con potencias

$$A^pA^q = A^{p+q},$$
$$(A^p)^q = A^{pq},$$

y además, si las matrices A y B son invertibles y $AB = BA$, se tiene

$$(AB)^p = A^pB^p.$$

Veamos ahora la relación que existe entre las operaciones de transposición e inversión. Aplicando a las relaciones (1) la regla de transposición del producto de matrices, obtenemos

$$X'A' = A'X' = E,$$

es decir, al transponer una matriz invertible A se obtiene de nuevo una matriz invertible y

$$(A')^{-1} = (A^{-1})'. \tag{5}$$

Una matriz cuadrada A se llama *ortogonal*, si

$$AA' = A'A = E, \tag{6}$$

es decir, si la matriz transpuesta es inversa de la inicial. De aquí se deduce, en particular, que *toda matriz ortogonal es invertible*.

Puesto que $(A')' = A$, de (6) se deduce que *la inversión de una matriz ortogonal es una matriz ortogonal*.

Además, si las matrices A y B son ortogonales, se tiene

$$A' = A^{-1}, \quad B' = B^{-1}$$

y, por consiguiente,

$$(AB)' = B'A' = B^{-1}A^{-1} = (AB)^{-1}.$$

En otras palabras, *el producto de matrices ortogonales es una matriz ortogonal*.

Consideremos una operación matricial más. Sea A una matriz arbitraria, cuyos elementos son números complejos. Sustituyamos

en A todo elemento por el número complejo conjugado. La matriz obtenida por este procedimiento se llama *conjugada compleja* de A y se designa por $\overline{A}$. La operación consistente en el paso a la matriz conjugada compleja posee las propiedades siguientes:

$$\overline{\alpha A + \beta B} = \bar{\alpha}\,\overline{A} + \bar{\beta}\,\overline{B},$$
$$\overline{AB} = \overline{A}\,\overline{B},$$
$$\overline{A'} = (\overline{A})',$$
$$\overline{A^{-1}} = (\overline{A})^{-1};$$

la demostración es muy sencilla y queda a cargo del lector.

Las matrices A y $\overline{A}'$ se denominan *conjugadas según Hermite* [1]). Si $A = \overline{A}'$, la matriz A se llama *hermitiana* o *simétrica según Hermite.*

Una matriz A que satisface la relación

$$\overline{A}'A = A\overline{A}' = E,$$

se llama *unitaria.*

Se puede demostrar, por el mismo procedimiento que el empleado para las matrices ortogonales, que *la matriz inversa de una matriz unitaria es unitaria* y que *el producto de matrices unitarias es también una matriz unitaria.*

Si todos los elementos de la matriz A son números reales, se tiene $\overline{A} = A$ y, por consiguiente, para las matrices reales los conceptos de simetría y de simetría según Hermite, así como los de matriz unitaria y de matriz ortogonal, coinciden.

1.4. Matrices celulares. Dividamos una matriz A en partes mediante un sistema de rectas verticales y horizontales. Estas partes pueden ser consideradas como matrices de órdenes inferiores que forman, interpretadas como elementos, la propia matriz; se denominan *células, cajas* o *bloques* de la matriz A, mientras que la propia matriz A, dividida de un modo determinado en células, se denomina, respectivamente, celular, de caja o de bloque. Una misma matriz puede ser dividida en células de diferentes maneras; por ejemplo:

$$\left[\begin{array}{ccc} 1 & 8 & 7 & 6 \\ 3 & 5 & 0 & 2 \\ 1 & 4 & 9 & 3 \end{array}\right], \quad \left[\begin{array}{ccc} 1 & 8 & 7 & 6 \\ 3 & 5 & 0 & 2 \\ 1 & 4 & 9 & 3 \end{array}\right], \quad \left[\begin{array}{ccc} 1 & 8 & 7 & 6 \\ 3 & 5 & 0 & 2 \\ 1 & 4 & 9 & 3 \end{array}\right].$$

La conveniencia de la división en células consiste en que las operaciones principales sobre matrices celulares se realizan formalmente siguiendo las mismas reglas que en el caso de matrices co-

[1]) En el capítulo V estas matrices se llaman también conjugadas transpuestas o anticonjugadas *(N. del Tr.)*

rrientes. En efecto, supongamos una matriz A dividida de algún modo en células:

$$A=\begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ \dots & \dots & \dots & \dots \\ A_{m1} & A_{m2} & \dots & A_{mn} \end{bmatrix}.$$

Al multiplicar todas las células por un número α multiplicaremos, al mismo tiempo, todos los elementos de la matriz A por α. Por consiguiente,

$$\alpha A=\begin{bmatrix} \alpha A_{11} & \alpha A_{12} & \dots & \alpha A_{1n} \\ \dots & \dots & \dots & \dots \\ \alpha A_{m1} & \alpha A_{m2} & \dots & \alpha A_{mn} \end{bmatrix}.$$

Sea B una matriz dividida en el mismo número de células que la matriz A:

$$B=\begin{bmatrix} B_{11} & B_{12} & \dots & B_{1n} \\ \dots & \dots & \dots & \dots \\ B_{m1} & B_{m2} & \dots & B_{mn} \end{bmatrix}.$$

Supongamos, además, que las correspondientes células de las matrices A y B son del mismo número de filas y de columnas respectivamente.

Para sumar las matrices A y B hay que sumar, según la definición, sus elementos correspondientes. Pero lo mismo ocurrirá, si sumamos las células correspondientes de estas matrices. Por esto

$$A+B=\begin{bmatrix} A_{11}+B_{11} & A_{12}+B_{12} & \dots & A_{1n}+B_{1n} \\ \dots & \dots & \dots & \dots \\ A_{m1}+B_{m1} & A_{m2}+B_{m2} & \dots & A_{mn}+B_{mn} \end{bmatrix}.$$

La situación es menos evidente en el caso de la multiplicación. Consideremos las matrices

$$U=\begin{bmatrix} U_{11} & \dots & U_{1n} \\ \dots & \dots & \dots \\ U_{m1} & & U_{mn} \end{bmatrix}, \qquad V=\begin{bmatrix} V_{11} & \dots & V_{1p} \\ \dots & \dots & \dots \\ V_{n1} & & V_{np} \end{bmatrix},$$

divididas en células U_{ij} y V_{jk} de manera que el número de columnas de la célula U_{ij} sea igual al número de filas de la célula V_{jk} ($i=1, \dots, m$; $j=1, \dots, n$; $k=1, \dots, p$). En estas condiciones las expresiones

$$W_{ik}=U_{i1}V_{1k}+U_{i2}V_{2k}+\dots+U_{in}V_{nk}$$

tienen sentido. Es fácil demostrar que

$$UV = \begin{bmatrix} W_{11} & \dots & W_{1p} \\ \dots & \dots & \dots \\ W_{m1} & \dots & W_{mp} \end{bmatrix}, \tag{1}$$

es decir, las matrices divididas de manera adecuada en células pueden ser multiplicadas de la forma corriente: *las células del producto son iguales a las sumas de los productos de las células de las filas de U por las células correspondientes de las columnas de V.*

Demostremos primero esta regla en el siguiente caso particular:

$$[AB]\begin{bmatrix} C \\ D \end{bmatrix} = AC + BD. \tag{2}$$

Sean α_{ij}, β_{ik}, γ_{jl} y δ_{kl} los elementos de las matrices A, B, C y D respectivamente ($i = 1, \dots, m$; $j = 1, \dots, n$; $k = 1, \dots, s$; $l = 1, \dots, t$). Efectuando las operaciones indicadas en el primer miembro de la igualdad (2), obtendremos que el elemento que se halla en la i-ésima fila y l-ésima columna es igual a

$$\alpha_{i1}\gamma_{1l} + \dots + \alpha_{in}\gamma_{nl} + \beta_{i1}\delta_{1l} + \dots + \beta_{is}\delta_{sl}.$$

Por otra parte, calculando el elemento correspondiente del segundo miembro, obtendremos la misma expresión y, por consiguiente, la igualdad (2) queda demostrada.

Empleando la fórmula (2), es fácil demostrar ahora una fórmula más general

$$[A_1 A_2 \dots A_n]\begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_n \end{bmatrix} = A_1B_1 + A_2B_2 + \dots + A_nB_n. \tag{3}$$

donde A_i y B_i son células. Para $n = 2$ esta fórmula coincide con (2). Apliquemos la inducción. Supongamos que para los valores de n menores que un valor dado la fórmula (3) ha sido ya demostrada y sea

$$C = [A_2 \dots A_n], \quad D = \begin{bmatrix} B_2 \\ \vdots \\ B_n \end{bmatrix}.$$

En este caso de (2) obtenemos

$$[A_1 A_2 \dots A_n]\begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_n \end{bmatrix} = [A_1 C]\begin{bmatrix} B_1 \\ D \end{bmatrix} = A_1B_1 + CD = A_1B_1 + A_2B_2 + \dots + A_nB_n.$$

De manera análoga se obtienen las fórmulas

$$A[B_1\ B_2\ \dots\ B_n] = [AB_1\ AB_2\ \dots\ AB_n], \tag{4}$$

$$\begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{bmatrix} B = \begin{bmatrix} A_1B \\ A_2B \\ \vdots \\ A_nB \end{bmatrix}. \tag{5}$$

Para deducir ahora de las fórmulas particulares (3), (4) y (5) la fórmula general (1), designemos mediante $U_1, \dots, U_m$ las filas de células de la matriz U y mediante

$V_1, \ldots, V_p$ las columnas de células de la matriz V. En base a la fórmula (5) tenemos

$$UV = \begin{bmatrix} U_1 \\ \vdots \\ U_m \end{bmatrix} V = \begin{bmatrix} U_1 V \\ \vdots \\ U_m V \end{bmatrix}.$$

Tomando aquí en lugar de la matriz V su división en células $[V_1 \ldots V_p]$ y empleando la fórmula (4), obtenemos

$$UV = \begin{bmatrix} U_1 V_1 & \ldots & U_1 V_p \\ \ldots & \ldots & \ldots \\ U_m V_1 & \ldots & U_m V_p \end{bmatrix}. \tag{6}$$

Por otra parte, en virtud de (3) tenemos

$$U_i V_k = [U_{i1} \ \ldots \ U_{in}] \begin{bmatrix} V_{1k} \\ \vdots \\ V_{nk} \end{bmatrix} = U_{i1} V_{1k} + \ldots + U_{in} V_{nk} = W_{ik}.$$

Colocando estas expresiones en (6), obtendremos la fórmula (1).

En el caso de matrices cuadradas resulta necesario, como regla general, dividirlas de manera que las células diagonales también sean cuadradas. Es fácil ver que, divididas dos matrices cuadradas en células de manera que las células diagonales sean cuadradas y que los ordenes de las células diagonales correspondientes coincidan, esta división satisface tanto las condiciones en las que es posible la adición célula por célula, como las condiciones que son necesarias para poder multiplicarlas como matrices celulares.

Toda matriz celular de tipo

$$A = \begin{bmatrix} A_1 & O & \ldots & O \\ O & A_2 & \ldots & O \\ \ldots & \ldots & \ldots & \ldots \\ O & O & \ldots & A_s \end{bmatrix},$$

donde $A_1, \ldots, A_s$ son células cuadradas y O son matrices nulas de dimensiones adecuadas, se llama matriz *celular diagonal*. En el mismo sentido se dice también que A se descompone en partes $A_1, \ldots, A_s$ o que A es *la suma directa* de las matrices $A_1, \ldots, A_s$; simbólicamente

$$A = A_1 \dot{+} A_2 \dot{+} \ldots \dot{+} A_s.$$

Las operaciones con matrices descompuestas se reducen a las operaciones con sus células diagonales. De aquí, a su vez, se desprende que siendo $f(\lambda)$ un polinomio y A una matriz celular

diagonal de células diagonales $A_1, \ldots, A_s$, se tiene

$$f(A) = \begin{bmatrix} f(A_1) & & & \\ & f(A_2) & & \\ & & \ddots & \\ & & & f(A_s) \end{bmatrix}. \tag{7}$$

1.5. Cuaternios. Las matrices constituyen un instrumento cómodo mediante el cual se pueden construir a partir de un anillo dado, por ejemplo, del anillo de los números reales, anillos de estructura más compleja. De forma sistemática este problema se estudia en la teoría de anillos y nos vamos a limitar a considerar solamente dos casos particulares.

Consideremos el anillo R_2 de todas las matrices cuadradas de segundo orden sobre el cuerpo R de los números reales. Tomemos

$$E = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \text{e} \quad I = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

Sea C el conjunto de todas las matrices de R_2 de tipo

$$\alpha E + \beta I = \begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix} \quad (\alpha, \beta \in R). \tag{1}$$

Elevando al cuadrado la matriz I, obtenemos $I^2 = -E$ y por ello las operaciones con las matrices (1) se pueden efectuar siguiendo las fórmulas

$$(\alpha E + \beta I) \pm (\gamma E + \delta I) = (\alpha \pm \gamma) E + (\beta \pm \delta) I,$$
$$(\alpha E + \beta I)(\gamma E + \delta I) = (\alpha\gamma - \beta\delta) E + (\alpha\delta + \beta\gamma) I,$$

es decir, las mismas fórmulas que para los números complejos $\alpha + \beta i$ y $\gamma + \delta i$. De las fórmulas señaladas se deduce también que el conjunto de matrices C es un anillo y que la correspondencia

$$\alpha E + \beta I \longrightarrow \alpha + \beta i$$

es una aplicación isomorfa del anillo C sobre el anillo de todos los números complejos.

A partir de las matrices de segundo orden E e I construimos ahora cuatro matrices cuadradas de orden 4:

$$e = \begin{bmatrix} E & 0 \\ 0 & E \end{bmatrix}, \quad i = \begin{bmatrix} 0 & -E \\ E & 0 \end{bmatrix}, \quad j = \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix}, \quad k = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}. \tag{2}$$

La matriz e es la matriz unidad corriente de orden 4 y, por ello,

$$ei = ie = i, \quad ej = je = j, \quad ek = ke = k. \tag{3}$$

Multiplicando las matrices (2) según las reglas de multiplicación de matrices celulares y teniendo en cuenta que $I^2 = -E$, obtenemos

$$ij = k = -ji, \quad jk = i = -kj, \quad ki = j = -ik \tag{4}$$

y, además,

$$i^2 = j^2 = k^2 = -e. \tag{5}$$

Es fácil memorizar las fórmulas (4): significan que en la secuencia i, j, k, i, j, k el producto de dos elementos consecutivos es igual al elemento que les sigue.

Las matrices de tipo

$$\alpha e + \beta i + \gamma j + \delta k = \begin{bmatrix} \alpha & -\gamma & -\beta & -\delta \\ \gamma & \alpha & \delta & -\beta \\ \beta & -\delta & \alpha & \gamma \\ \delta & \beta & -\gamma & \alpha \end{bmatrix}, \tag{6}$$

donde α, β, γ, δ son números reales arbitrarios, se llaman *cuaternios* (o, a veces, matrices cuaternios). De (6) se deduce que la representación de los cuaternios en la forma $\alpha e + \beta i + \gamma j + \delta k$ es unívoca. En otras palabras, la relación

$$\alpha e + \beta i + \gamma j + \delta k = \alpha_1 e + \beta_1 i + \gamma_1 j + \delta_1 k$$

equivale a cuatro igualdades:

$$\alpha = \alpha_1, \quad \beta = \beta_1, \quad \gamma = \gamma_1, \quad \delta = \delta_1.$$

La adición y la sustracción de los cuaternios, representados en la forma algebraica normal $\alpha e + \beta i + \gamma j + \delta k$, se realiza por la regla corriente:

$$\begin{aligned} &(\alpha e + \beta i + \gamma j + \delta k) \pm (\alpha_1 e + \beta_1 i + \gamma_1 j + \delta_1 k) = \\ &= (\alpha \pm \alpha_1) e + (\beta \pm \beta_1) i + (\gamma \pm \gamma_1) j + (\delta \pm \delta_1) k. \end{aligned} \tag{7}$$

Para multiplicar dos cuaternios, representados en la forma algebraica normal, es suficiente recurrir a la ley distributiva y a las tablas de multiplicación (3), (4) y (5). Como resultado llegamos a la fórmula

$$\begin{aligned} &(\alpha e + \beta i + \gamma j + \delta k)(\alpha_1 e + \beta_1 i + \gamma_1 j + \delta_1 k) = \\ &= (\alpha\alpha_1 - \beta\beta_1 - \gamma\gamma_1 - \delta\delta_1) e + (\alpha\beta_1 + \beta\alpha_1 + \gamma\delta_1 - \delta\gamma_1) i + \\ &+ (\alpha\gamma_1 + \gamma\alpha_1 + \delta\beta_1 - \beta\delta_1) j + (\alpha\delta_1 + \delta\alpha_1 + \beta\gamma_1 - \gamma\beta_1) k. \end{aligned} \tag{8}$$

Las fórmulas (7) y (8) muestran que el conjunto Q de los cuaternios matriciales es un anillo con unidad e, que constituye un subanillo del anillo de todas las matrices reales de orden 4. Los cuaternios e, i, j y k suelen llamarse cuaternios unidades. De las relaciones (4) se deduce que el anillo de los cuaternios es no conmutativo. El hecho más notable consiste en que el anillo de

los cuaternios es un cuerpo, es decir, que en el anillo de los cuaternios pueden ser resueltas todas las ecuaciones de tipo

$$ax = b, \quad ya = b, \tag{9}$$

donde a y b son cualesquiera dos cuaternios dados, de los cuales a es diferente de cero. Más abajo se dan unas fórmulas cómodas para la solución de estas ecuaciones.

Consideremos un cuaternio cualquiera

$$q = \alpha e + \beta i + \gamma j + \delta k.$$

El número real

$$N(q) = \alpha^2 + \beta^2 + \gamma^2 + \delta^2$$

se llama *norma* del cuaternio q. Puesto que α, β, γ y δ son números reales, la norma de un cuaternio es un número real no negativo e igual a cero sólo para el cuaternio nulo.

El cuaternio

$$q^* = \alpha e - \beta i - \gamma j - \delta k$$

se denomina *conjugado* de q. Es claro que $q^{**} = q$. Mediante la multiplicación directa de los cuaternios q y q^* (según la fórmula (8)) se obtienen las igualdades principales

$$qq^* = q^*q = N(q) \cdot e,$$

de donde

$$q \cdot \frac{1}{N(q)} \cdot q^* = e, \tag{10}$$

o

$$q^{-1} = \frac{1}{N(q)} q^* \tag{11}$$

(siempre que $q \neq 0$).

Volvamos ahora a las ecuaciones (9). Multiplicándolas por a^{-1}, la primera por la izquierda y la segunda por la derecha, obtenemos

$$x = a^{-1}b, \quad y = ba^{-1}. \tag{12}$$

Con la sustitución de estos valores en las ecuaciones (9) se demuestra que las fórmulas (12) ofrecen efectivamente las soluciones buscadas de estas ecuaciones.

En el caso general las soluciones $a^{-1}b$ y ba^{-1} son distintas. Por esto suelen llamarse cocientes por la izquierda y por la derecha de b por a, designándose mediante $a \backslash b$ y b / a, respectivamente. Efectuando el cálculo directo, es fácil demostrar las fórmulas

$$(\alpha a + \beta b)^* = \alpha a^* + \beta b^*,$$
$$(ab)^* = b^* a^*,$$

de donde es fácil deducir la importante relación

$$N(ab) = N(a)\,N(b) \quad (a,\ b \in Q).$$

En efecto,

$$N(ab)\cdot e = abb^*a^* = aa^*N(b) = N(a)\,N(b)\cdot e.$$

La aplicación $\alpha \longrightarrow \alpha e$ es un isomorfismo del cuerpo (conmutativo) de los números reales en el cuerpo de los cuaternios. Esto permite identificar un cuaternio de tipo αe con el número α y en lugar de $\alpha e + \beta i + \gamma j + \delta k$ escribir simplemente $\alpha + \beta i + \gamma j + \delta k$. Así, por ejemplo, se tiene

$$(j+k)(1+i-j)^{-1} = (j+k)(1-i+j)\frac{1}{3} = -\frac{1}{3} - \frac{1}{3}i + \frac{2}{3}k.$$

La mayor parte de los resultados de las secciones siguientes del libro estará relacionada con objetos cuya definición depende de un cuerpo K dado de antemano. Aunque tendrán interés principal los casos en que K es un cuerpo conmutativo, los razonamientos se realizarán de manera que no quede excluido el caso de cuerpos no conmutativos. Al analizar estos razonamientos conviene tener en cuenta que el cuerpo K puede ser precisamente el cuerpo de los cuaternios Q.

Ejemplos y problemas

1. Sean

$$\varphi(\lambda) = -2 - 5\lambda + 3\lambda^2 \quad \text{y} \quad A = \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix};$$

entonces

$$\varphi(A) = \begin{bmatrix} 14 & 2 \\ 3 & 14 \end{bmatrix}.$$

2. Demuéstrese que la matriz

$$U = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}$$

es unitaria. ¿Bajo qué condiciones una matriz diagonal resulta ser ortogonal? ¿unitaria?

3. Hállese la inversa de la matriz

$$A = \begin{bmatrix} 2 & 2 & 1 \\ 3 & 1 & 5 \\ 3 & 2 & 3 \end{bmatrix}.$$

4. Si

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix},$$

se tiene

$$A^n = \begin{bmatrix} 1 & n & \frac{1}{2}n(n-1) \\ 0 & 1 & n \\ 0 & 0 & 1 \end{bmatrix}.$$

5. Demuéstrese que todas las matrices conmutables con

$$A=\begin{bmatrix}0&1&0&0\\0&0&1&0\\0&0&0&1\\0&0&0&0\end{bmatrix}$$

son de tipo

$$B=\begin{bmatrix}\alpha&\beta&\gamma&\delta\\0&\alpha&\beta&\gamma\\0&0&\alpha&\beta\\0&0&0&\alpha\end{bmatrix}.$$

6. Si una matriz posee dos de las tres propiedades siguientes: es *real*, es *ortogonal*, es *unitaria*, posee también la tercera.

7. Toda matriz cuadrada puede ser representada como la suma de una matriz simétrica y otra antisimétrica.

8. Una matriz I se llama *involutiva*, si $I^2=E$. Demuéstrese que si una matriz posee dos de las propiedades: es *simétrica*, es *ortogonal*, es *involutiva*, posee también la tercera.

9. Una matriz P se llama *idempotente*, si $P^2=P$. Demuéstrese que las matrices

$$\begin{bmatrix}-26&-18&-27\\21&15&21\\12&8&13\end{bmatrix} \text{ y } \begin{bmatrix}1&0&0\\0&1&0\\0&0&0\end{bmatrix}$$

son idempotentes.

10. Si P es idempotente, la matriz

$$I=2P-E$$

es involutiva y viceversa, si I es involutiva, la matriz

$$P=\frac{1}{2}(I+E)$$

es idempotente.

11. Consideremos las matrices cuadradas de orden n. Sea F_{ij} ($i, j=1, \ldots, \ldots, n$) la matriz en la que el elemento de la i-ésima fila y j-ésima columna es igual a 1, mientras que los demás elementos son iguales a 0. En estas condiciones, para $A=\|\alpha_{ij}\|$ se tiene

$$A\cdot E_{ij}=\alpha_{1i}E_{1j}+\ldots+\alpha_{ni}E_{nj},$$
$$E_{ij}\cdot A=\alpha_{j1}E_{i1}+\ldots+\alpha_{jn}E_{in}.$$

Dedúzcase de aquí que la matriz A es conmutable con cada una de las matrices E_{ij} si, y sólo si, A es de la forma αE.

Empleando este resultado, demuéstrese que la matriz A es conmutable con una matriz cualquiera cuadrada de orden n si, y sólo si, $A=\alpha E$, donde α es un elemento del anillo principal K conmutable con cualquier elemento de K.

12. A veces, además de matrices de orden finito, se consideran también matrices de orden infinito que tienen la forma de una tabla infinita de dos entradas:

$$A=\begin{bmatrix}\alpha_{11}&\alpha_{12}&\ldots\\\alpha_{21}&\alpha_{22}&\ldots\\\ldots&\ldots&\ldots\end{bmatrix}.$$

Una matriz de este tipo se llama *finita en fila*, si cada una de sus filas contiene solamente un número finito de elementos diferentes de cero. Las operaciones con las matrices finitas en fila (así como con las matrices finitas en columna, que se definen análogamente) se realizan siguiendo las mismas reglas que tienen lugar en el caso de matrices cuadradas de orden finito. Es fácil ver que el resultado es de nuevo una matriz finita en fila.

Sea $A = \|\alpha_{ij}\|$ una matriz de orden infinito tal que $1 = \alpha_{12} = \alpha_{23} = \ldots$ y $\alpha_{st} = 0$ para $t - s \neq 1$. Demuéstrese que $AA' = E$ y $A'A \neq E$.

§ 2. Determinantes

2.1. Definición. El concepto de determinante surgió en relación con el problema de solución de sistemas de ecuaciones lineales. Tomemos un cuerpo conmutativo K y consideremos sistemas elementales de ecuaciones de primer grado de dos y tres incógnitas y con coeficientes de K. Un sistema de dos ecuaciones lineales con dos incógnitas ξ_1 y ξ_2 se representa en la forma siguiente:

$$\begin{aligned} \alpha_{11}\xi_1 + \alpha_{12}\xi_2 &= \beta_1, \\ \alpha_{21}\xi_1 + \alpha_{22}\xi_2 &= \beta_2, \end{aligned} \tag{1}$$

donde α_{ij} y β_i son números de K dados. Las matrices

$$A = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \quad \text{y} \quad B = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \beta_1 \\ \alpha_{21} & \alpha_{22} & \beta_2 \end{bmatrix}$$

se llaman matriz *principal* y matriz *ampliada*, respectivamente, del sistema (1). Con el fin de eliminar la incógnita ξ_2, multipliquemos la primera ecuación por α_{22} y la segunda por $-\alpha_{12}$ y sumemos ambas. Obtendremos la ecuación

$$(\alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21})\xi_1 = \beta_1\alpha_{22} - \beta_2\alpha_{12}.$$

Si $\alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21} \neq 0$, obtenemos de esta ecuación y de una ecuación análoga que se obtiene eliminando ξ_1

$$\xi_1 = \frac{\alpha_{22}\beta_1 - \alpha_{12}\beta_2}{\alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21}}, \qquad \xi_2 = \frac{\alpha_{11}\beta_2 - \alpha_{21}\beta_1}{\alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21}}. \tag{2}$$

Los denominadores de las expresiones de las incógnitas ξ_1 y ξ_2 coinciden y representan un polinomio en los elementos de la matriz principal A. El valor de este polinomio se llama *determinante* de la matriz A y se designa $\det A$ o $|A|$. Si la matriz viene dada por su tabla, el determinante se designa escribiendo la tabla entre barras verticales.

Es decir, por definición para cualquier matriz cuadrada de orden dos se tiene

$$\begin{vmatrix} \alpha & \beta \\ \gamma & \delta \end{vmatrix} = \alpha\delta - \beta\gamma. \tag{3}$$

Empleando los determinantes, la fórmula (2) puede ser escrita en la forma

$$\xi_1 = \frac{\begin{vmatrix} \beta_1 & \alpha_{12} \\ \beta_2 & \alpha_{22} \end{vmatrix}}{\begin{vmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{vmatrix}}, \qquad \xi_2 = \frac{\begin{vmatrix} \alpha_{11} & \beta_1 \\ \alpha_{21} & \beta_2 \end{vmatrix}}{\begin{vmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{vmatrix}}. \tag{4}$$

Resolviendo de forma análoga el sistema de tres ecuaciones

$$\begin{aligned} \alpha_{11}\xi_1 + \alpha_{12}\xi_2 + \alpha_{13}\xi_3 &= \beta_1, \\ \alpha_{21}\xi_1 + \alpha_{22}\xi_2 + \alpha_{23}\xi_3 &= \beta_2, \\ \alpha_{31}\xi_1 + \alpha_{32}\xi_2 + \alpha_{33}\xi_3 &= \beta_3 \end{aligned} \tag{5}$$

con tres incógnitas ξ_1, ξ_2 y ξ_3, obtenemos

$$\xi_1 = \frac{\beta_1\alpha_{22}\alpha_{33} - \beta_1\alpha_{23}\alpha_{32} + \beta_2\alpha_{23}\alpha_{12} - \beta_2\alpha_{12}\alpha_{33} + \beta_3\alpha_{12}\alpha_{23} - \beta_3\alpha_{22}\alpha_{13}}{\alpha_{11}\alpha_{22}\alpha_{33} - \alpha_{11}\alpha_{23}\alpha_{32} + \alpha_{12}\alpha_{23}\alpha_{31} - \alpha_{12}\alpha_{21}\alpha_{33} + \alpha_{13}\alpha_{21}\alpha_{32} - \alpha_{13}\alpha_{22}\alpha_{31}} \tag{6}$$

y unas expresiones análogas para ξ_2 y ξ_3. Por supuesto estas expresiones tienen sentido sólo en el caso en que su denominador sea diferente de cero. Las matrices

$$A = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{bmatrix} \quad \text{y} \quad B = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \beta_1 \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \beta_2 \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \beta_3 \end{bmatrix}$$

también se llaman matriz *principal* y matriz *ampliada* del sistema de ecuaciones (5). El denominador de la fórmula (6) se llama *determinante* de la matriz cuadrada A de orden tres. Luego, por definición,

$$\begin{vmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{vmatrix} = \alpha_{11}\alpha_{22}\alpha_{33} - \alpha_{11}\alpha_{23}\alpha_{32} + \alpha_{12}\alpha_{23}\alpha_{31} - \alpha_{12}\alpha_{21}\alpha_{33} + \alpha_{13}\alpha_{21}\alpha_{32} - \alpha_{13}\alpha_{22}\alpha_{31}. \tag{7}$$

Uniendo en el segundo miembro los términos que contienen α_{11}, α_{12} y α_{13} y recordando la fórmula (3), obtenemos

$$\begin{vmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{vmatrix} = \alpha_{11}\begin{vmatrix} \alpha_{22} & \alpha_{23} \\ \alpha_{32} & \alpha_{33} \end{vmatrix} - \alpha_{12}\begin{vmatrix} \alpha_{21} & \alpha_{23} \\ \alpha_{31} & \alpha_{33} \end{vmatrix} + \alpha_{13}\begin{vmatrix} \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{vmatrix}. \tag{8}$$

Es fácil memorizar la fórmula (8). Para abreviar, en lugar de *determinante de una matriz* de orden dos o tres suele decirse *determinante* de segundo o tercer orden. Los tres determinantes de segundo orden de la fórmula (8) se obtienen suprimiendo del determinante del orden tres, que figura en la misma fórmula, la primera fila y, respectivamente, la primera, la segunda y la tercera columna. A continuación, el determinante de segundo orden que se obtiene suprimiendo la primera fila y la j-ésima columna debe ser multi-

plicado por el elemento que se halla en la primera fila y en la j-ésima columna y los productos obtenidos deben ser tomados con signos alternos y sumados. Como resultado obtendremos el determinante de orden tres.

Esta regla nos sugiere la idea de cómo debe ser definido el concepto de determinante de una matriz cuadrada de orden cuatro, cinco y de órdenes superiores. Por lo tanto, introducimos la siguiente definición principal:

DEFINICIÓN. *Se llama determinante de una matriz de primer orden, formada por el número α, el propio número α. Supongamos ahora que para un número natural cualquiera $n \geqslant 1$ conocemos ya el elemento del anillo K que representa el determinante de una matriz arbitraria cuadrada de orden n sobre K. Entonces, para una matriz cuadrada arbitraria $A = \|\alpha_{ij}\|$ de orden $n+1$ sobre K tomamos, por definición,*

$$\det \|\alpha_{ij}\| = \alpha_{11} |A_1^1| - \alpha_{12} |A_1^2| + \alpha_{13} |A_1^3| - \ldots + (-1)^n \alpha_{1, n+1} |A_1^{n+1}|, \tag{9}$$

donde $|A_1^j|$ es el valor del determinante de la matriz de orden n que se obtiene de la matriz inicial A suprimiendo la primera fila y la j-ésima columna ($j = 1, \ldots, n+1$).

Aplicando esta definición en el caso en que $n = 1$, obtenemos la fórmula (3) para determinantes de orden dos. Conociendo la expresión para los determinantes de orden dos, podemos emplear la definición principal para obtener la expresión (8) para los determinantes de orden tres. De la expresión (8) obtenemos mediante (3) la fórmula definitiva (7).

Veamos ahora cuál es la fórmula definitiva para los determinantes de orden cuatro. De acuerdo con la definición principal, el determinante de una matriz cuadrada arbitraria $A = \|\alpha_{ij}\|$ de orden cuatro coincide con la expresión

$$\begin{aligned} \alpha_{11} \begin{vmatrix} \alpha_{22} & \alpha_{23} & \alpha_{24} \\ \alpha_{32} & \alpha_{33} & \alpha_{34} \\ \alpha_{42} & \alpha_{43} & \alpha_{44} \end{vmatrix} - \alpha_{12} \begin{vmatrix} \alpha_{21} & \alpha_{23} & \alpha_{24} \\ \alpha_{31} & \alpha_{33} & \alpha_{34} \\ \alpha_{41} & \alpha_{43} & \alpha_{44} \end{vmatrix} + \alpha_{13} \begin{vmatrix} \alpha_{21} & \alpha_{22} & \alpha_{24} \\ \alpha_{31} & \alpha_{32} & \alpha_{34} \\ \alpha_{41} & \alpha_{42} & \alpha_{44} \end{vmatrix} - \\ - \alpha_{14} \begin{vmatrix} \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \\ \alpha_{41} & \alpha_{42} & \alpha_{43} \end{vmatrix} \end{aligned} \tag{10}$$

Introduciendo aquí las expresiones de los determinantes de orden tres según la fórmula (7) y suprimiendo los paréntesis, obtendremos la fórmula definitiva que buscabamos para un determinante de orden cuatro. No la escribiremos puesto que no tiene sentido memorizarla. Según (7) todo determinante de tercer orden es igual a la suma de seis términos tomados con signos alternados. Por eso, si

tomamos en (10) en lugar de los determinantes de orden tres sus expresiones y suprimimos los paréntesis, obtendremos en total $4 \cdot 6 = 24$ términos. La mitad de ellos aparecerá con el signo más y la otra mitad con el signo menos. Una afirmación análoga es válida también para los determinantes de orden cualquiera.

TEOREMA 1 *Para una matriz cuadrada arbitraria se tiene*

$$\begin{vmatrix} \alpha_{11} & \dots & \alpha_{1n} \\ \dots & \dots & \dots \\ \alpha_{n1} & \dots & \alpha_{nn} \end{vmatrix} = \sum \pm \alpha_{1i_1} \alpha_{2i_2} \dots \alpha_{ni_n}, \tag{11}$$

donde la suma se extiende a las permutaciones arbitrarias $(i_1, i_2, \dots, i_n)$ *de los números* $1, 2, \dots, n$. *El signo más o menos se toma según sea par o impar la permutación* $(i_1, i_2, \dots, i_n)$, *es decir, en la mitad de los casos se toma el signo más y en la otra mitad se toma el signo menos; el número total de términos en la suma* (11) *es igual a* $1 \cdot 2 \dots n = n!$.

Para $n = 1, 2, 3, 4$ la veracidad del teorema ya ha sido comprobada. Ahora aplicamos la inducción. Supongamos que el teorema es cierto para los determinantes de un orden cualquiera n y sea $A = \|\alpha_{ij}\|$ una matriz cuadrada de orden $n+1$. Por hipótesis, cada determinante $|A_1^l|$ $(l = 1, \dots, n+1)$ de la fórmula (9) es de la forma:

$$|A_1^l| = \sum \pm \alpha_{2m_1} \alpha_{3m_2} \dots \alpha_{n+1, m_n}, \tag{12}$$

donde la suma se extiende a todas las permutaciones $(m_1, \dots, m_n)$ de los números $1, \dots, l-1, l+1, \dots, n+1$. El número de términos en la suma (12) es igual a $n!$. Sustituyendo en (9) $|A_1^l|$ por sus expresiones y suprimiendo los paréntesis, obtendremos un total de $n!(n+1) = (n+1)!$ términos. La mitad de ellos tendrá el signo más y la otra mitad el signo menos. No habrá términos semejantes, puesto que los términos que se obtienen al suprimir distintos paréntesis difieren en el primer factor. Es evidente que el término arbitrario de tipo $\alpha_{1i_1} \alpha_{2i_2} \dots \alpha_{n+1, i_{n+1}}$ se obtiene al suprimir los paréntesis en el producto $\alpha_{1i_1} |A_1^{i_1}|$ y, por consiguiente, la fórmula (11) es verídica.

De la fórmula (11) se puede deducir fácilmente el siguiente corolario importante. Supongamos que los elementos de la matriz $A = \|\alpha_{ij}\|$ son números complejos. En virtud de la fórmula (11), tenemos

$$|\overline{A}| = \sum \pm \bar{\alpha}_{1i_1} \dots \bar{\alpha}_{ni_n} = \overline{\sum \pm \alpha_{1i_1} \dots \alpha_{ni_n}} = \overline{|A|},$$

es decir,

$$|\overline{A}| = \overline{|A|}.$$

Según la fórmula (11) el valor del determinante de una matriz es igual a una suma algebraica de términos, cada uno de los cuales es el producto de elementos, tomados de manera que haya uno de

cada fila y uno de cada columna. Por ello, si todos los elementos de una fila o de una columna de la matriz son iguales a cero, también serán iguales a cero todos los términos del determinante. Es decir, hemos obtenido el siguiente corolario:

COROLARIO. *Si una fila o una columna de una matriz cuadrada está compuesta íntegramente por ceros, su determinante es igual a cero.*

Una matriz cuadrada se llama *semidescompuesta*, si sus elementos pueden ser divididos mediante una línea vertical y otra horizontal en cuatro matrices de manera que a lo largo de la diagonal figuren matrices cuadradas y una de las otras dos matrices esté compuesta íntegramente por ceros. En otras palabras, la matriz A es semidescompuesta si tiene una de las dos formas siguientes

$$\begin{bmatrix} \alpha_{11} & \dots & \alpha_{1r} & \alpha_{1,r+1} & \dots & \alpha_{1n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \alpha_{r1} & \dots & \alpha_{rr} & \alpha_{r,r+1} & \dots & \alpha_{rn} \\ 0 & \dots & 0 & \alpha_{r+1,r+1} & \dots & \alpha_{r+1,n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & \alpha_{n,r+1} & \dots & \alpha_{nn} \end{bmatrix},$$

$$\begin{bmatrix} \alpha_{11} & \dots & \alpha_{1r} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \alpha_{r1} & \dots & \alpha_{rr} & 0 & \dots & 0 \\ \alpha_{r+1,1} & \dots & \alpha_{r+1,r} & \alpha_{r+1,r+1} & \dots & \alpha_{r+1,n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \alpha_{n1} & \dots & \alpha_{nr} & \alpha_{n,r+1} & \dots & \alpha_{nn} \end{bmatrix}.$$

A veces, la primera de estas matrices suele llamarse matriz semidescompuesta *superior* y la segunda, matriz semidescompuesta *inferior*.

TEOREMA 2. *El determinante de una matriz semidescompuesta es igual al producto de los determinantes de sus células diagonales.*

Para una matriz de orden dos esta proposición es evidente, ya que

$$\begin{vmatrix} \alpha & \beta \\ 0 & \delta \end{vmatrix} = \begin{vmatrix} \alpha & 0 \\ \gamma & \delta \end{vmatrix} = \alpha\delta.$$

Apliquemos ahora la inducción aceptando que el teorema 2 es cierto para las matrices de un orden cualquiera $n-1$. Consideremos una matriz arbitraria semidescompuesta A de orden n. Supongamos que es de la forma

$$A = \|\alpha_{ij}\| = \begin{bmatrix} B & D \\ O_{sr} & C \end{bmatrix},$$

donde B y C son matrices cuadradas de orden r y s respectivamente ($r+s=n$) mientras que O_{sr} y D son matrices rectangulares y la matriz O_{sr} es nula. Aplicando la fórmula (9), obtenemos

$$|A| = \alpha_{11}|A_1^1| - \alpha_{12}|A_1^2| + \dots + (-1)^{n-1}\alpha_{1n}|A_1^n|, \qquad (13)$$

donde $|A_1^i|$ es la matriz de orden $n-1$ que se obtiene suprimiendo en A la primera fila y la i-ésima columna. Es fácil ver que todas las matrices A_1^i son semidescompuestas y, por ello, según la suposición inductiva, el determinante de cada una de ellas es igual al producto de los determinantes de las células diagonales. Pero las matrices $A_1^1, \ldots, A_1^r$ y las matrices $A_1^{r+1}, \ldots, A_1^n$ se descomponen de modos distintos. Para las primeras tenemos

$$|A_1^i| = |B_1^i| \cdot |C|, \quad i = 1, \ldots, r, \tag{14}$$

mientras que las matrices $A_1^{r+1}, \ldots, A_1^n$ pueden ser divididas mediante una línea vertical y otra horizontal de manera que el cuadrado superior de la izquierda sea de orden r. Puesto que la última fila de este cuadrado está compuesta de ceros, su determinante es igual a cero. El determinante de la matriz A_1^{r+i} es igual al producto del determinante del cuadrado indicado por el determinante del cuadrado complementario. Luego, $|A_1^{r+i}| = 0$ y, por consiguiente,

$$\begin{aligned} |A| &= \alpha_{11} |B_1^1| \cdot |C| - \ldots + (-1)^{r-1} \alpha_{1r} |B_1^r| \cdot |C| = \\ &= (\alpha_{11} |B_1^1| - \ldots + (-1)^{r-1} \alpha_{1r} |B_1^r|) \cdot |C| = |B| \cdot |C|, \end{aligned} \tag{15}$$

que es lo que queríamos demostrar. El caso en que la matriz A es de la forma

$$A = \|\alpha_{ij}\| = \begin{bmatrix} B & O_{rs} \\ D & C \end{bmatrix},$$

es aun más sencillo, ya que ahora se tiene $\alpha_{1, r+1} = \ldots = \alpha_{1n} = 0$ y por esto de (13) y (14) obtenemos inmediatamente (15).

TEOREMA 3 *Si en las matrices cuadradas $A = \|\alpha_{ij}\|$ y $B = \|\beta_{ij}\|$ de un mismo orden n coinciden todos los elementos correspondientes menos los elementos de una fila i-ésima cualquiera, se tiene*

$$\begin{vmatrix} \alpha_{11} & \ldots & \alpha_{1n} \\ \ldots & \ldots & \ldots \\ \alpha_{i1} & \ldots & \alpha_{in} \\ \ldots & \ldots & \ldots \\ \alpha_{n1} & \ldots & \alpha_{nn} \end{vmatrix} + \begin{vmatrix} \alpha_{11} & \ldots & \alpha_{1n} \\ \ldots & \ldots & \ldots \\ \beta_{i1} & \ldots & \beta_{in} \\ \ldots & \ldots & \ldots \\ \alpha_{n1} & \ldots & \alpha_{nn} \end{vmatrix} = \begin{vmatrix} \alpha_{11} & \ldots & \alpha_{1n} \\ \ldots & \ldots & \ldots \\ \alpha_{i1} + \beta_{i1} & \ldots & \alpha_{in} + \beta_{in} \\ \ldots & \ldots & \ldots \\ \alpha_{n1} & \ldots & \alpha_{nn} \end{vmatrix}. \tag{16}$$

Este teorema suele llamarse a veces teorema de adición de determinantes. Pasando a su demostración, designemos mediante C la matriz del segundo miembro de la igualdad (16). Para matrices de primer orden la igualdad (16) es evidente. Aplicando la inducción, supongamos que para matrices de orden $n-1$ el teorema 3 es cierto. Si en la fórmula (16) $i = 1$, desarrollando el determinante de la matriz C según la fórmula (9), obtenemos

$$|C| = (\alpha_{11} + \beta_{11}) |C_1^1| - \ldots + (-1)^{n-1} (\alpha_{1n} + \beta_{1n}) |C_1^n|.$$

Es evidente que $C_1^1 = A_1^1 = B_1^1$ y, por consiguiente, se tiene

$$|C| = (\alpha_{11}|A_1^1| - \ldots + (-1)^{n-1}\alpha_{1n}|A_1^n|) + \\ + (\beta_{11}|B_1^1| - \ldots + (-1)^{n-1}\beta_{1n}|B_1^n|) = |A| + |B|.$$

Supongamos ahora que en la fórmula (16) $i > 1$. Entonces, se tiene

$$|C| = \alpha_{11}|C_1^1| - \ldots + (-1)^{n-1}\alpha_{1n}|C_1^n|. \qquad (17)$$

Las matrices $C_1^1, \ldots, C_1^n$ son de orden $n-1$ y por esto para ellas es válido el desarrollo de tipo (16). Como resultado, obtenemos

$$|C_1^j| = |A_1^j| + |B_1^j| \quad (j = 1, \ldots, n)$$

y, en virtud de la relación (17), tenemos $|C| = |A| + |B|$.

Desde el punto de vista formal, la definición principal de determinante sirve también para matrices formadas por elementos de un anillo asociativo K cualquiera (no necesariamente conmutativo). En la demostración de los teoremas 1, 2 y 3 tampoco se ha empleado la conmutatividad de K. Sin embargo, una serie de propiedades de los determinantes, de importancia para las aplicaciones, dependen de la conmutatividad del anillo principal K. Una de estas propiedades se indica en el siguiente teorema.

TEOREMA 4. *Sea A una matriz cuadrada formada por elementos de un anillo conmutativo K. Si se multiplican todos los elementos de una fila de la matriz A por un elemento $\lambda \in K$, el determinante de la matriz también se multiplicará por λ. En otras palabras, se tiene*

$$\begin{vmatrix} \alpha_{11} & \ldots & \alpha_{1n} \\ \ldots & \ldots & \ldots \\ \lambda\alpha_{i1} & \ldots & \lambda\alpha_{in} \\ \ldots & \ldots & \ldots \\ \alpha_{n1} & \ldots & \alpha_{nn} \end{vmatrix} = \lambda \begin{vmatrix} \alpha_{11} & \ldots & \alpha_{1n} \\ \ldots & \ldots & \ldots \\ \alpha_{i1} & \ldots & \alpha_{in} \\ \ldots & \ldots & \ldots \\ \alpha_{n1} & \ldots & \alpha_{nn} \end{vmatrix}.$$

La demostración se realiza igual que en el teorema anterior y no la vamos a repetir. Señalemos solamente el siguiente corolario importante:

COROLARIO. *Para cualquier matriz cuadrada A de orden n formada por elementos de un anillo conmutativo K y para cualquier $\lambda \in K$ se tiene*

$$|\lambda A| = \lambda^n |A|.$$

En efecto, al multiplicar la matriz A por λ, cada una de sus n filas se multiplicará por λ. Por consiguiente, el determinante de la matriz A se multiplicará por λ^n.

2.2. Propiedades principales de los determinantes. Consideremos una matriz cuadrada arbitraria $A = \|\alpha_{ij}\|$ de orden n formada por

elementos de un anillo cualquiera K. Según la definición principal, se tiene

$$|A| = \alpha_{11}|A_1^1| - \alpha_{12}|A_1^2| + \ldots + (-1)^{n+1}\alpha_{1n}|A_1^n|. \qquad (1)$$

donde A_i^j es la matriz que se obtiene de A suprimiendo la primera fila y la j-ésima columna. Aplicando ahora al determinante de la matriz A_i^j la misma fórmula, obtenemos

$$|A_i^j| = \alpha_{21}|A_{12}^{1j}| - \ldots + (-1)^{j}\alpha_{2,\,j-1}|A_{12}^{1,\,j-1}| + \\ + (-1)^{j+1}\alpha_{2,\,j+1}|A_{12}^{1,\,j+1}| + \ldots + (-1)^{n}\alpha_{2n}|A_{12}^{1n}|, \qquad (2)$$

donde A_{12}^{jk} es la matriz que se obtiene de A suprimiendo la primera y la segunda filas y la j-ésima y la k-ésima columnas ($j \neq k$).

Introduzcamos en la fórmula (1) en lugar de los valores $|A_1^j|$ sus valores (2). Como resultado, llegamos a la relación

$$|A| = \sum_{j \neq k} \pm \alpha_{1j}\alpha_{2k}|A_{12}^{jk}|. \qquad (3)$$

Aquí la suma se extiende a todos los pares posibles j, k de diferentes números, pertenecientes al conjunto $1, \ldots, n$ y el signo más o menos se toma de acuerdo con las fórmulas (1) y (2). Las matrices A_{12}^{jk} y A_{12}^{kj} coinciden evidentemente y, por lo tanto, la fórmula (3) puede ser representada en la forma siguiente:

$$|A| = \sum_{j<k} (\pm \alpha_{1j}\alpha_{2k} \pm \alpha_{1k}\alpha_{2j})|A_{12}^{jk}|. \qquad (4)$$

Calculemos ahora con mayor exactitud cuáles son los signos que deben tomarse en la última fórmula. En virtud de (2), para el j-ésimo término de (1) tenemos

$$(-1)^{j+1}\alpha_{1j}|A_1^j| = \ldots + (-1)^{j+1}\alpha_{1j}(-1)^{k}\alpha_{2k}|A_{12}^{jk}| + \ldots$$

Análogamente y teniendo en cuenta que $j < k$, obtenemos para el k-ésimo término de (1)

$$(-1)^{k+1}\alpha_{1k}|A_1^k| = \ldots + (-1)^{k+1}\alpha_{1k}(-1)^{j+1}\alpha_{2j}|A_{12}^{jk}| + \ldots$$

Por consiguiente, el coeficiente del término $|A_{12}^{jk}|$ de la fórmula (4) es igual a

$$(-1)^{j+k+1}(\alpha_{1j}\alpha_{2k} - \alpha_{1k}\alpha_{2j}) = (-1)^{j+k+1+2}\begin{vmatrix}\alpha_{1j} & \alpha_{1k}\\ \alpha_{2j} & \alpha_{2k}\end{vmatrix}$$

y, en consecuencia,

$$|A| = \sum_{j<k} (-1)^{j+k+1+2}\begin{vmatrix}\alpha_{1j} & \alpha_{1k}\\ \alpha_{2j} & \alpha_{2k}\end{vmatrix} \cdot |A_{12}^{jk}|. \qquad (5)$$

La fórmula (5) se denomina desarrollo del determinante según los elementos de la primera y segunda filas. Es fácil memorizarla:

se toman todos los determinantes posibles de orden dos formados por los elementos que se hallan en la primera y segunda filas y en las columnas j-ésima y k-ésima ($j < k$) y se multiplican por los determinantes de las matrices correspondientes que se obtienen al suprimir en la matriz A las filas y columnas indicadas. Los productos se multiplican además por $(-1)^{j+k+1+2}$, donde el exponente es igual a la suma de los números que corresponden a las filas y a las columnas suprimidas, y después se suman. La suma algebraica así obtenida es igual al determinante de la matriz dada.

Reglas semejantes tienen lugar también para los desarrollos según las tres primeras, las cuatro primeras, etc. filas. Pero en lo sucesivo no las necesitaremos.

Emplearemos ahora la fórmula (5) para deducir una serie de propiedades principales de los determinantes. En lo que sigue se supone que los elementos de las matrices consideradas se toman de un anillo conmutativo K.

TEOREMA 1. *Si en una matriz cuadrada se cambian entre sí dos filas cualesquiera, el determinante de la matriz nueva será igual al determinante de la matriz inicial tomado con el signo menos.*

Para matrices de segundo orden esta proposición se comprueba directamente:

$$\begin{vmatrix} \alpha & \beta \\ \gamma & \delta \end{vmatrix} = \alpha\delta - \beta\gamma, \qquad \begin{vmatrix} \gamma & \delta \\ \alpha & \beta \end{vmatrix} = \gamma\beta - \delta\alpha = -(\alpha\delta - \beta\gamma).$$

A continuación, suponemos por inducción que el teorema es justo para las matrices de orden $n-1$ y que la matriz dada $A = \|\alpha_{ij}\|$ es de orden n. Supongamos que en A se cambian entre sí las dos primeras filas. De la fórmula (5) vemos que todos los determinantes de orden dos cambian de signo, mientras que los factores adicionales no varían. Luego, toda la suma adquiere el factor -1 que es lo que queríamos demostrar.

Consideremos el caso en el que se cambian entre sí la j-ésima y la k-ésima filas, donde $1 < j < k$. Entonces, la primera fila permanece invariable y del desarrollo (1) deducimos que cada factor $|A_1^j|$ obtendrá después del intercambio el valor opuesto; por ello, toda la suma obtendrá después del intercambio de las filas el valor opuesto.

Finalmente, supongamos que se cambian entre sí la primera y la i-ésima fila, donde $i > 2$. Este mismo resultado obtendremos si cambiamos entre sí primero la primera y la segunda filas, luego la segunda y la i-ésima y, finalmente, la segunda y la primera. Según hemos demostrado el determinante cambiará cada vez de signo y después de los tres cambios el determinante se multiplicará por -1.

COROLARIO 1. *El determinante de una matriz cuadrada en la que coinciden dos filas es igual a cero.*

Cambiando las filas entre sí se puede conseguir que las fillas coincidentes sean la primera y la segunda. El determinante de la matriz con las filas cambiadas o bien coincidirá con el determinante de la matriz inicial o bien diferirá de él en el factor — 1. De la fórmula (5) se ve directamente que el determinante de una matriz en la que coinciden las dos primeras filas es igual a cero. Por esto será también igual a cero el determinante de la matriz inicial.

COROLARIO 2. *Si a los elementos de la r-ésima fila cualquiera de la matriz cuadrada* $A = \|\alpha_{ij}\|$ *se agregan los elementos correspondientes de otra fila s-ésima cualquiera* ($s \neq r$) *multiplicados por un factor arbitrario* λ, *el determinante de la matriz nueva será igual al determinante de la inicial.*

En efecto, según los teoremas 3 y 4 del punto anterior, se tiene

$$\begin{vmatrix} \alpha_{11} & \dots & \alpha_{1n} \\ \dots & \dots & \dots \\ \alpha_{r1} & \dots & \alpha_{rn} \\ \dots & \dots & \dots \\ \alpha_{s1}+\lambda\alpha_{r1} & \dots & \alpha_{sn}+\lambda\alpha_{rn} \\ \dots & \dots & \dots \\ \alpha_{n1} & \dots & \alpha_{nn} \end{vmatrix} = \begin{vmatrix} \alpha_{11} & \dots & \alpha_{1n} \\ \dots & \dots & \dots \\ \alpha_{r1} & \dots & \alpha_{rn} \\ \dots & \dots & \dots \\ \alpha_{s1} & \dots & \alpha_{sn} \\ \dots & \dots & \dots \\ \alpha_{n1} & \dots & \alpha_{nn} \end{vmatrix} + \lambda \begin{vmatrix} \alpha_{11} & \dots & \alpha_{1n} \\ \dots & \dots & \dots \\ \alpha_{r1} & \dots & \alpha_{rn} \\ \dots & \dots & \dots \\ \alpha_{r1} & \dots & \alpha_{rn} \\ \dots & \dots & \dots \\ \alpha_{n1} & \dots & \alpha_{nn} \end{vmatrix}.$$

Aquí el último determinante tiene dos filas coincidentes y, por consiguiente, es igual a cero.

Hasta el momento todos nuestros resultados estaban relacionados con las filas de los determinantes. Hagamos el primer paso para introducir en el juego las columnas.

TEOREMA 2. *Para toda matriz cuadrada* $A = \|\alpha_{ij}\|$ *de orden n es válido el siguiente desarrollo según los elementos de la primera columna:*

$$|A| = \alpha_{11}|A_1^1| - \alpha_{21}|A_2^1| + \dots + (-1)^{n-1}\alpha_{n1}|A_n^1|, \qquad (6)$$

donde A_r^1 *es la matriz que se obtiene de A suprimiendo la primera columna y la r-ésima fila.*

Consideremos que los elementos α_{ij} de la matriz dada son letras. Entonces, el determinante de la matriz A será un polinomio en estas letras cuya forma general se ha establecido en el teorema 1 del punto anterior. En particular, hemos señalado allí que cada término del polinomio $|A|$ contiene un factor, y sólo uno, perteneciente al conjunto $\alpha_{11}, \alpha_{21}, \dots, \alpha_{n1}$. Agrupemos en $|A|$ todos los términos que contienen el factor α_{r1}, saquemos fuera de los paréntesis este factor común y designemos mediante A_{r1} la expresión comprendida entre los paréntesis. De esta forma obtendremos

$$|A| = \alpha_{11}A_{11} + \alpha_{21}A_{21} + \dots + \alpha_{n1}A_{n1}. \qquad (7)$$

Comparando las fórmulas (7) y (1), llegamos a la conclusión de que

$$A_{11} = |A_1^1|.$$

Para hallar la expresión análoga para A_{r1}, es suficiente recurrir ahora al teorema 1. En efecto, cambiemos sucesivamente la r-ésima fila de la matriz A con cada una de las anteriores, elevándola más y más. Después de $r-1$ intercambios de esta índole obtendremos una matriz B que difiere de la matriz A sólo en el orden de filas y, por consiguiente, tendremos

$$|A| = (-1)^{r-1} |B|.$$

Desarrollando el determinante B según los elementos de la primera fila, obtenemos

$$|B| = \alpha_{r1} |B_1^1| - \alpha_{r2} |B_1^2| + \ldots + (-1)^{n-1} \alpha_{rn} |B_1^n|,$$

y por lo tanto

$$|A| = \alpha_{r1} (-1)^{r-1} |B_1^1| - \ldots + (-1)^{r+n-2} \alpha_{rn} |B_1^n|.$$

Comparando este desarrollo con el de la fórmula (7), llegamos a la relación

$$A_{r1} = (-1)^{r-1} |B_1^1|.$$

La matriz B_1^1 se obtiene suprimiendo en la matriz B la primera fila y la primera columna. Está claro que la misma matriz se obtendrá al suprimir en A la r-ésima fila y la primera columna, es decir, que $B_1^1 = A_r^1$. Sustituyendo en la fórmula (7) el valor A_{r1} por el valor $(-1)^{r-1} |A_r^1|$, obtenemos el desarrollo deseado (6).

Anteriormente hemos indicado que la matriz, en la que la primera, la segunda, etc. filas coinciden respectivamente con la primera, la segunda, etc. columnas de la matriz A, se llama *transpuesta* respecto de A y se designa por A'. Es evidente, que siendo A una matriz cuadrada de orden n, la matriz A' es también una matriz cuadrada de orden n.

TEOREMA 3. *El determinante de una matriz cuadrada no varía en la transposición, es decir,*

$$|A'| = |A|. \tag{8}$$

Para matrices de orden uno la proposición es evidente. Recurriendo, al igual que antes, a la inducción, aceptemos que la matriz dada $A = \|\alpha_{ij}\|$ es de orden $n > 1$ y que el teorema es cierto para las matrices de orden $n-1$. Desarrollando el determinante de la matriz A según los elementos de la primera fila y el determinante de la matriz transpuesta A' según los elementos de la primera columna, obtenemos

$$\begin{aligned} |A| &= \alpha_{11} |A_1^1| - \alpha_{12} |A_1^2| + \ldots + (-1)^{n-1} \alpha_{1n} |A_1^n|, \\ |A'| &= \alpha_{11} |(A')_1^1| - \alpha_{12} |(A')_2^1| + \ldots + (-1)^{n-1} \alpha_{1n} |(A')_n^1|. \end{aligned} \tag{9}$$

Sin embargo, es fácil ver que $(A')_i' = (A_i^i)'$. Las matrices A_i^i son de orden $n-1$ y, por la hipótesis de inducción, se tiene

$$|(A')_i'| = |(A_i^i)'| = |A_i^i| \qquad (i = 1, \ldots, n).$$

Comparando estas relaciones con los desarrollos (9) obtenemos (8).

Por consiguiente, al calcular el determinante de una matriz las columnas y las filas pueden ser sustituidas unas por otras. Esto significa que de todo teorema referente a las propiedades del determinante de una matriz, enunciado en términos de filas o de columnas, se puede obtener un nuevo teorema sustituyendo las filas y las columnas unas por otras. En particular, de las propiedades de los determinantes indicadas anteriormente y relacionadas con las filas, obtenemos el resultado siguiente.

COROLARIO. *Al cambiar entre sí dos columnas de una matriz su determinante adquiere el valor opuesto. El determinante de una matriz cuadrada que tiene dos columnas idénticas es igual a cero. Si se multiplican todos los elementos de una columna de la matriz por* λ, *el determinante de la matriz también quedará multiplicado por* λ. *Si se agregan a todos los elementos de una columna de la matriz los elementos correspondientes de otra columna, multiplicados por un número fijo, el determinante de la matriz nueva será igual al determinante de la inicial.*

Análogamente, sustituyendo las filas por las columnas obtenemos del desarrollo (5) el desarrollo según las dos primeras columnas:

$$|A| = \sum_{i<j} (-1)^{i+j+1+s} \begin{vmatrix} a_{i1} & a_{is} \\ a_{j1} & a_{js} \end{vmatrix} \cdot |A_{ij}^{1s}|.$$

Finalmente, del teorema de adición de determinantes, enunciado en términos de filas, obtenemos la fórmula correspondiente

$$\begin{vmatrix} \alpha_{11} & \ldots & \alpha_{1i} & \ldots & \alpha_{1n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \alpha_{n1} & \ldots & \alpha_{ni} & \ldots & \alpha_{nn} \end{vmatrix} + \begin{vmatrix} \alpha_{11} & \ldots & \beta_{1i} & \ldots & \alpha_{1n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \alpha_{n1} & \ldots & \beta_{ni} & \ldots & \alpha_{nn} \end{vmatrix} =$$

$$= \begin{vmatrix} \alpha_{11} & \ldots & \alpha_{1i} + \beta_{1i} & \ldots & \alpha_{1n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \alpha_{n1} & \ldots & \alpha_{ni} + \beta_{ni} & \ldots & \alpha_{nn} \end{vmatrix}$$

referente a las columnas.

Hasta el momento nos hemos valido de los desarrollos de un determinante según los elementos de su *primera* fila o su *primera* columna. Al mismo tiempo, conocemos la ley de variación del determinante al cambiar en él entre sí filas o columnas. Esto ofrece la posibilidad de obtener inmediatamente de los desarrollos según los elementos de la primera fila o la primera columna los desarrollos análogos según los elementos de cualquier fila o columna.

TEOREMA 4. *Para una matriz cuadrada arbitraria* $A = \|\alpha_{ij}\|$ *de orden n son válidos los siguientes desarrollos según los elementos de la r-ésima fila y s-ésima columna:*

$$|A| = (-1)^{r+1}\alpha_{r1}|A_r^1| + \ldots + (-1)^{r+n}\alpha_{rn}|A_r^n|, \qquad (10)$$

$$|A| = (-1)^{s+1}\alpha_{1s}|A_1^s| + \ldots + (-1)^{s+n}\alpha_{ns}|A_n^s|, \qquad (11)$$

donde A_i^j *es la matriz que se obtiene de A suprimiendo la i-ésima fila y la j-ésima columna.*

Puesto que las filas y las columnas se encuentran en las mismas condiciones basta demostrar sólo una de las fórmulas (10) y (11), por ejemplo, la fórmula (10). Cambiando sucesivamente la r-ésima fila de la matriz A con cada una de las anteriores, después de $r-1$ intercambios obtendremos la matriz

$$B = \begin{vmatrix} \alpha_{r1} & \alpha_{r2} & & \alpha_{rn} \\ \alpha_{11} & \alpha_{12} & \ldots & \alpha_{1n} \\ \ldots & \ldots & \ldots & \ldots \\ \alpha_{r-1,1} & \alpha_{r-1,2} & \ldots & \alpha_{r-1,n} \\ \alpha_{r+1,1} & \alpha_{r+1,2} & \ldots & \alpha_{r+1,n} \\ \ldots & \ldots & \ldots & \ldots \\ \alpha_{n1} & \alpha_{n2} & \ldots & \alpha_{nn} \end{vmatrix}.$$

Según el teorema sobre el intercambio de filas tenemos $|A| = (-1)^{r-1}|B|$. Desarrollando aquí el determinante B según los elementos de la primera fila y empleando las relaciones evidentes $B_1^j = A_r^j$, llegamos al desarrollo (10).

El determinante de la matriz A_i^j se llama *menor* del determinante de la matriz A correspondiente al elemento α_{ij}. La expresión $(-1)^{i+j}|A_i^j|$ se llama *adjunto* del elemento α_{ij} en $|A|$ y se designa con frecuencia mediante $|A|_{ij}$. Empleando el concepto de adjunto, las fórmulas (10) y (11) pueden ser representadas en forma más breve:

$$\alpha_{i1}|A|_{i1} + \alpha_{i2}|A|_{i2} + \ldots + \alpha_{in}|A|_{in} = |A|, \qquad (12)$$

$$\alpha_{1i}|A|_{1i} + \alpha_{2i}|A|_{2i} + \ldots + \alpha_{ni}|A|_{ni} = |A|. \qquad (13)$$

En estas igualdades cada elemento α_{ij} se multiplica por su adjunto $|A|_{ij}$. ¿Qué sucederá si se toma la suma de los productos de los elementos de la i-ésima fila por los adjuntos de los elementos correspondientes de otra fila cualquiera? Con el fin de obtener la respuesta, sustituyamos la j-ésima fila de la matriz dada A por su i-ésima fila sin alterar todas las filas restantes, incluyendo también la i-ésima. Obtendremos una matriz B en la que son idénticas las filas i-ésima y j-ésima. El determinante de esta matriz es igual a cero. Al mismo tiempo, es evidente que los menores de los elementos de la j-ésima fila en los determinantes de las matrices

A y B coinciden. Desarrollando B según los elementos de la i-ésima fila, obtenemos

$$\alpha_{i1}|A|_{j1}+\ldots+\alpha_{in}|A|_{jn}=0 \qquad (i\neq j). \tag{14}$$

Sustituyendo en los razonamientos las filas por las columnas obtenemos la segunda serie de relaciones:

$$\alpha_{1i}|A|_{1j}+\ldots+\alpha_{ni}|A|_{nj}=0 \qquad (i\neq j). \tag{15}$$

Las relaciones (12), (13), (14) y (15) suelen enunciarse de la forma siguiente.

TEOREMA 5. *La suma de los productos de los elementos de una fila (de una columna) de un determinante por sus adjuntos es igual al valor del determinante. La suma de los productos de los elementos de una fila (de una columna) de un determinante por los adjuntos de los elementos correspondientes de otra fila (columna) cualquiera es igual a cero.*

Hagamos ahora algunas observaciones acerca de los métodos de cálculo de determinantes. Los determinantes de orden dos y tres se calculan generalmente mediante las fórmulas iniciales. Por ejemplo,

$$\begin{vmatrix} 7 & -2 & 5 \\ 2 & 4 & 1 \\ 3 & -1 & 8 \end{vmatrix} = 7\begin{vmatrix} 4 & 1 \\ -1 & 8 \end{vmatrix} + 2\begin{vmatrix} 2 & 1 \\ 3 & 8 \end{vmatrix} + 5\begin{vmatrix} 2 & 4 \\ 3 & -1 \end{vmatrix} = 187.$$

Aquí hemos desarrollado el determinante de orden tres según los elementos de su primera fila. Si uno de los elementos del determinante dado fuese igual a cero, resultaría más conveniente recurrir al desarrollo según aquella fila (o columna) que contenga este elemento nulo. Este mismo método se puede aplicar al calcular determinantes de orden elevado que contienen muchos elementos iguales a cero. En particular, tenemos

$$\begin{vmatrix} \alpha_{11} & \alpha_{12} & \ldots & \alpha_{1n} \\ 0 & \alpha_{22} & \ldots & \alpha_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \alpha_{nn} \end{vmatrix} = \alpha_{11}\begin{vmatrix} \alpha_{22} & \ldots & \alpha_{2n} \\ \ldots & \ldots & \ldots \\ 0 & \ldots & \alpha_{nn} \end{vmatrix} = \alpha_{11}\alpha_{22}\ldots\alpha_{nn}$$

y análogamente

$$\begin{vmatrix} 0 & 0 & \ldots & \alpha_{1n} \\ \ldots & \ldots & \ldots & \ldots \\ 0 & \alpha_{n-1,2} & \ldots & \alpha_{n-1,n} \\ \alpha_{n1} & \alpha_{n2} & \ldots & \alpha_{nn} \end{vmatrix} = (-1)^{n+1}\alpha_{n1}\begin{vmatrix} 0 & \ldots & \alpha_{1n} \\ \ldots & \ldots & \ldots \\ \alpha_{n-1,2} & \ldots & \alpha_{n-1,n} \end{vmatrix} =$$

$$= (-1)^{\frac{n(n-1)}{2}}\alpha_{n1}\ldots\alpha_{1n}.$$

La situación es más compleja si el orden del determinante dado es relativamente elevado (por ejemplo, 7 y más) y entre los elementos del determinante hay pocos ceros o no los hay. En este caso, al determinante se aplican primero las transformaciones indicadas en el corolario 2 del teorema 1, tratando de escoger el coeficiente λ de modo que el determinante nuevo contenga en algunos lugares ceros o tenga en cierto sentido una estructura más sencilla. Estos métodos se ven con claridad en el ejemplo siguiente. Supongamos que debemos calcular el determinante

$$d=\begin{vmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 3 & 4 & 5 & 6 & 1 \\ 3 & 4 & 5 & 6 & 1 & 2 \\ 4 & 5 & 6 & 1 & 2 & 3 \\ 5 & 6 & 1 & 2 & 3 & 4 \\ 6 & 1 & 2 & 3 & 4 & 5 \end{vmatrix}.$$

Restando en este determinante la quinta fila de la sexta, restando en el determinante obtenido la cuarta fila de la quinta, restando después la tercera fila de la cuarta, etc., obtenemos

$$d=\begin{vmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 1 & 1 & 1 & 1 & -5 \\ 1 & 1 & 1 & 1 & -5 & 1 \\ 1 & 1 & 1 & -5 & 1 & 1 \\ 1 & 1 & -5 & 1 & 1 & 1 \\ 1 & -5 & 1 & 1 & 1 & 1 \end{vmatrix}.$$

Restando ahora la sexta fila de la primera, de la segunda, ..., de la quinta, obtenemos

$$d=\begin{vmatrix} 0 & 7 & 2 & 3 & 4 & 5 \\ 0 & 6 & 0 & 0 & 0 & -6 \\ 0 & 6 & 0 & 0 & -6 & 0 \\ 0 & 6 & 0 & -6 & 0 & 0 \\ 0 & 6 & -6 & 0 & 0 & 0 \\ 1 & -5 & 1 & 1 & 1 & 1 \end{vmatrix}=(-1)6^4\begin{vmatrix} 7 & 2 & 3 & 4 & 5 \\ 1 & 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & -1 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \end{vmatrix}.$$

En el determinante obtenido de orden cinco agregamos a la primera columna las restantes y después desarrollamos el determinante según los elementos de la primera columna. Tendremos

$$d=-6^4\cdot\begin{vmatrix} 21 & 2 & 3 & 4 & 5 \\ 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \end{vmatrix}=-6^4\cdot 21.$$

Mediante cálculos análogos se obtiene también la fórmula general

$$\begin{vmatrix} 1 & 2 & 3 & \dots & n-1 & n \\ 2 & 3 & 4 & \dots & n & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ n & 1 & 2 & \dots & n-2 & n-1 \end{vmatrix} = (-1)^{\frac{n(n-1)}{2}} \frac{n(n-1)}{2} n^{n-2}. \qquad (16)$$

Este determinante lleva el nombre de *circulador* de orden n, debido a que sus filas se obtienen por permutaciones cíclicas de los elementos de la primera fila. Es esta regularidad en la disposición de los elementos del determinante en la que se basa la deducción de la fórmula (16) que, a propósito, no será empleada en lo sucesivo. Actualmente para hallar el valor numérico de los determinantes de orden elevado se recurre a los servicios de los centros de cómputo en los que existen programas standard para el cálculo de determinantes adaptados a aquellos tipos de máquinas con las que está equipado el centro. Hallemos la estimación aproximada del número de operaciones aritméticas suficiente indudablemente para el cálculo del determinante de una matriz arbitraria $\|\alpha_{ij}\|$ de orden n.

Supongamos que $\alpha_{11} \neq 0$. Calculamos α_{11}^{-1} y representamos el determinante dado en la forma

$$\alpha_{11} \cdot \begin{vmatrix} 1 & \alpha_{11}^{-1}\alpha_{12} & \dots & \alpha_{11}^{-1}\alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2n} \\ \dots & \dots & \dots & \dots \\ \alpha_{n1} & \alpha_{n2} & \dots & \alpha_{nn} \end{vmatrix}.$$

Durante esta operación realizamos una inversión y $n-1$ multiplicaciones. A continuación, agregamos a los elementos de la segunda, de la tercera, ..., de la n-ésima filas los elementos de la primera fila multiplicados respectivamente por $-\alpha_{21}$, $-\alpha_{31}$, ..., $-\alpha_{n1}$ (un total de $(n-1)^2$ multiplicaciones e igual número de adiciones). Después de esto el problema se reduce al cálculo de un determinante de orden $n-1$:

$$\alpha_{11} \cdot \begin{vmatrix} \alpha'_{11} & \dots & \alpha_{1,\,n-1} \\ \dots & \dots & \dots \\ \alpha_{n-1,1} & \dots & \alpha_{n-1,\,n-1} \end{vmatrix}.$$

Continuando este proceso, tendremos que calcular al final del mismo solamente el producto $\alpha_{11}\alpha'_{11} \dots \alpha_{11}^{(n-1)}$ (es decir, $n-1$ multiplicaciones). Al aplicar el algoritmo expuesto necesitaremos realizar un total de $n-1$ inversiones de números, de

$$n(n-1)+(n-1)(n-2)+\dots+2\cdot 1 = \frac{(n+1)n(n-1)}{3}$$

multiplicaciones, de $\frac{(2n-1)n(n-1)}{6}$ adiciones y de $n-1$ multiplicaciones finales. Por lo tanto, para calcular un determinante de

orden n es necesario efectuar aproximadamente cerca de $\frac{1}{3}n^3$ multiplicaciones y de un número igual de adiciones. No nos detendremos aquí en el estudio más detallado de esta cuestión. Varias definiciones y resultados exactos se indican en el artículo de *В. Я. Пан*, „О способах вычисления значений многочленов", Успехи математических наук 21, № 1 (1966), 103—134 (V. Ya. Pan, Acerca de los métodos de cálculo de valores de polinomios), así como en la bibliografía señalada en este artículo.

2.3. Determinante de un producto. Matrices inversas. Un papel importante lo desempeña en la teoría de matrices el siguiente teorema.

TEOREMA 1 (sobre la multiplicación de determinantes). *El determinante del producto de matrices cuadradas (formadas por elementos pertenecientes a un anillo conmutativo K) es igual al producto de los determinantes de las matrices.*

Sean dadas dos matrices cuadradas $A=\|\alpha_{ij}\|$ y $B=\|\beta_{ij}\|$ de orden n. En virtud del teorema sobre las matrices semidescompuestas (p. 2.1), se tiene

$$|A|\cdot|B|=\begin{vmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1n} & 0 & 0 & \dots & 0 \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2n} & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \alpha_{n1} & \alpha_{n2} & \dots & \alpha_{nn} & 0 & 0 & \dots & 0 \\ -1 & 0 & \dots & 0 & \beta_{11} & \beta_{12} & \dots & \beta_{1n} \\ 0 & -1 & \dots & 0 & \beta_{21} & \beta_{22} & \dots & \beta_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & -1 & \beta_{n1} & \beta_{n2} & \dots & \beta_{nn} \end{vmatrix}.$$

En el segundo miembro aparece el determinante de una matriz de tipo especial de orden $2n$. Sin alterar el valor del determinante, con esta matriz se pueden realizar sucesivamente las siguientes transformaciones: agregar a su primera fila los elementos de la $(n+1)$-ésima fila multiplicados por α_{11}, los elementos de la $(n+2)$-ésima fila multiplicados por α_{12}, etc., los elementos de la $2n$-ésima fila multiplicados por α_{1n}. Así surgirá una matriz de orden $2n$, en la que las primeras n posiciones de la primera fila estarán ocupadas por ceros, mientras que las otras n posiciones estarán ocupadas por los productos de la primera fila de la matriz A por las columnas de la matriz B. En la nueva matriz de orden $2n$ agregamos ahora a los elementos de su segunda fila los elementos de la $(n+1)$-ésima fila, de la $(n+2)$-ésima fila, etc., multiplicados respectivamente por α_{21}, α_{22}, ..., α_{2n}. Después realizamos transformaciones análogas con la tercera, ..., con la n-ésima filas.

Después de esto obtenemos la igualdad siguiente:

$$|A|\cdot|B| = \begin{vmatrix} 0 \ldots 0 & \sum\alpha_{1i}\beta_{i1} & \ldots & \sum\alpha_{1i}\beta_{in} \\ \ldots\ldots\ldots & & & \\ 0 \ldots 0 & \sum\alpha_{ni}\beta_{i1} & \ldots & \sum\alpha_{ni}\beta_{in} \\ -1 \ldots 0 & \beta_{11} & \ldots & \beta_{1n} \\ \ldots\ldots\ldots & & & \\ 0 \ldots -1 & \beta_{n1} & \ldots & \beta_{nn} \end{vmatrix}.$$

Para reducir el determinante de la derecha a la forma semidescompuesta, cambiamos en él la primera columna con la $(n+1)$-ésima, la segunda con la $(n+2)$-ésima, ..., la n-ésima con la $2n$-ésima. Así obtendremos la igualdad

$$|A|\cdot|B| = (-1)^n \begin{vmatrix} \sum\alpha_{1i}\beta_{i1} & \ldots & \sum\alpha_{1i}\beta_{in} & 0 \ldots 0 \\ \ldots\ldots\ldots & & & \\ \sum\alpha_{ni}\beta_{i1} & \ldots & \sum\alpha_{ni}\beta_{in} & 0 \ldots 0 \\ \beta_{11} & \ldots & \beta_{1n} & -1 \ldots 0 \\ \ldots\ldots\ldots & & & \\ \beta_{n1} & \ldots & \beta_{nn} & 0 \ldots -1 \end{vmatrix}. \qquad (1)$$

Puesto que el determinante de una matriz semidescompuesta es igual al producto de los determinantes de las células diagonales, de la relación (1) se deduce que

$$|A|\cdot|B| = (-1)^n(-1)^n \begin{vmatrix} \sum\alpha_{1i}\beta_{i1} & \ldots & \sum\alpha_{1i}\beta_{in} \\ \ldots\ldots\ldots & & \\ \sum\alpha_{ni}\beta_{i1} & \ldots & \sum\alpha_{ni}\beta_{in} \end{vmatrix},$$

es decir,

$$|A|\cdot|B| = |AB|, \qquad (2)$$

que es lo que queríamos demostrar.

Hemos demostrado el teorema 1 para el producto de dos matrices. Está claro que de aquí se desprende su validez también en el caso del producto de un número finito cualquiera de matrices. Por ejemplo,

$$|ABC| = |(AB)C| = |AB|\cdot|C| = |A|\cdot|B|\cdot|C|.$$

En particular, para cualquier matriz cuadrada A se tiene

$$|A^k| = |A|^k \quad (k = 0, 1, 2, \ldots). \qquad (3)$$

Sabemos que la transposición de una matriz cuadrada no altera su determinante. Por ello, para dos matrices cuadradas cualesquiera de un mismo orden, se tiene

$$|A|\cdot|B| = |AB| = |A'B| = |AB'| = |A'B'|.$$

Sea, por ejemplo,

$$A = \begin{vmatrix} x & y \\ u & v \end{vmatrix}.$$

En este caso, se tiene

$$|A|^2 = |AA'| = \begin{vmatrix} x^2+y^2 & xu+yv \\ xu+yv & u^2+v^2 \end{vmatrix} = (x^2+y^2)(u^2+v^2) - (xu+yv)^2,$$

es decir,

$$(xv-yu)^2 = (x^2+y^2)(u^2+v^2) - (xu+yv)^2.$$

Esta relación suele escribirse en forma de la siguiente identidad

$$(x^2+y^2)(u^2+v^2) = (xv-yu)^2 + (xu+yv)^2$$

que se denomina *identidad de Lagrange*.

Empleemos ahora el teorema 1 para un estudio más detallado de las propiedades de la matriz inversa.

Consideremos una matriz cuadrada cualquiera $A = \|\alpha_{ij}\|$ de orden n. Formemos una matriz con los adjuntos de los elementos del determinante A y tomemos su transpuesta. La matriz así obtenida se llama *adjunta* de A y se designa mediante A^*. En otras palabras,

$$A^* = \begin{bmatrix} |A|_{11} & |A|_{21} & \dots & |A|_{n1} \\ |A|_{12} & |A|_{22} & \dots & |A|_{n2} \\ \dots & \dots & \dots & \dots \\ |A|_{1n} & |A|_{2n} & \dots & |A|_{nn} \end{bmatrix}.$$

Calculando los productos $A \cdot A^*$ y $A^* \cdot A$ según la regla de multiplicación de matrices y teniendo en cuenta las fórmulas de (12) a (15) del punto anterior, obtenemos directamente las relaciones importantes

$$A \cdot A^* = A^* \cdot A = |A| \cdot E, \tag{4}$$

donde E es la matriz unidad.

El determinante de la matriz A es un elemento del anillo principal K, del cual se toman los elementos de todas las matrices consideradas. Supongamos que para el elemento $|A|$ existe en K su inverso $|A|^{-1}$. Entonces, multiplicando las relaciones (4) por $|A|^{-1}$, llegamos a las igualdades

$$A \cdot |A|^{-1} A^* = |A|^{-1} A^* \cdot A = E,$$

de donde se deduce que

$$A^{-1} = |A|^{-1} \cdot A^*. \tag{5}$$

Por otra parte, si la ecuación matricial

$$AX = E$$

tiene solución, obtenemos pasando a los determinantes

$$|A| \cdot |X| = 1,$$

es decir, el elemento $|A|$ es invertible en el anillo K.

Uniendo los resultados obtenidos, llegamos a la proposición siguiente.

TEOREMA 2. *Una matriz cuadrada A, formada por elementos de un anillo conmutativo K con el elemento unidad, posee la matriz inversa, también formada por elementos de K, si, y sólo si, el determinante de A es invertible en el anillo K. Si la matriz inversa existe, su determinante es igual al inverso del valor del determinante de la matriz dada.*

Por ejemplo, en el anillo de todos los números enteros $K = \{0, \pm 1, \ldots\}$ poseen inverso solamente los elementos 1 y -1. Por lo tanto, *una matriz cuadrada formada por números enteros posee la matriz inversa, también formada por números enteros, si, y sólo si, el determinante de la matriz dada es igual a* ± 1.

Supongamos que se consideran matrices formadas por elementos de un cuerpo conmutativo. En un cuerpo conmutativo, todo elemento no nulo posee su inverso. Por ello, sobre un cuerpo conmutativo son invertibles aquellas matrices, y sólo aquéllas, cuyos determinantes son diferentes de cero.

Una matriz cuadrada cuyo determinante es igual a cero se llama *singular*. En el caso contrario, la matriz se llama *regular*. Por esto, la condición indicada anteriormente para que una matriz sea invertible puede ser enunciada de la forma siguiente: *una matriz cuadrada sobre un cuerpo conmutativo posee inversa si, y sólo si, es regular*.

Hemos visto que para matrices A regulares sobre el cuerpo conmutativo K, se tiene

$$|A^{-1}| = |A|^{-1}.$$

De aquí obtenemos

$$|A^{-m}| = |A^{-1} \ldots A^{-1}| = |A|^{-m} \qquad (m = 1, 2, \ldots),$$

de modo que la fórmula (3) es válida no sólo para los valores naturales de k, sino también para cualesquiera valores enteros de k (siempre que la matriz A sea invertible).

Según el p.1.3, una matriz cuadrada A se llama ortogonal, si

$$AA' = A'A = E.$$

Pasando a los determinantes, obtenemos $|A|^2 = 1$ y, por ello, si las matrices se consideran sobre un cuerpo conmutativo K, se tiene $|A| = \pm 1$. Gracias a esto todas las matrices ortogonales se descomponen en dos conjuntos de matrices: las matrices *ortogonales propias* de determinante $+1$ y las matrices *ortogonales impropias* de determinante -1. Todas las matrices ortogonales propias for-

man, respecto a la operación de multiplicación, un grupo que constituye un subgrupo del grupo de todas las matrices ortogonales.

Recordemos que una matriz cuadrada A definida sobre el cuerpo de los números complejos se llama unitaria si $A\overline{A}' = E$. Tomando los determinantes, obtenemos $|A| \cdot \overline{|A|} = 1$, es decir, *el módulo del determinante de una matriz unitaria es igual a la unidad.*

Para concluir hallemos el determinante de la matriz adjunta. Aceptemos que el anillo K es el anillo de los polinomios en letras x_{ij} $(i, j = 1, \ldots, n)$ con coeficientes enteros y consideremos la matriz $X = \|x_{ij}\|$. Según la fórmula (4) se tiene

$$X \cdot X^* = |X| \cdot E$$

y por lo tanto

$$|X| \cdot |X^*| = |X|^n.$$

Puesto que el polinomio $|X|$ es diferente de cero, la igualdad puede ser dividida por $|X|$ y así obtenemos la fórmula deseada

$$|X^*| = |X|^{n-1}, \tag{6}$$

ambos miembros de la cual son polinomios en las variables x_{ij} con coeficientes enteros. Pero, si coinciden dos polinomios de este tipo, también coinciden sus valores para cualesquiera valores de las variables pertenecientes a un anillo conmutativo arbitrario. Por esto, la fórmula (6) es válida para matrices cuadradas arbitrarias sobre cualquier anillo conmutativo.

2.4. Sistemas cramerianos de ecuaciones lineales. Consideremos el conjunto de condiciones

$$\left.\begin{aligned} \alpha_{11}\xi_1 + \ldots + \alpha_{1n}\xi_n &= \beta_1, \\ \cdots\cdots\cdots\cdots\cdots & \\ \alpha_{m1}\xi_1 + \ldots + \alpha_{mn}\xi_n &= \beta_m, \end{aligned}\right\} \tag{1}$$

donde $\xi_1, \ldots, \xi_n$ son letras, mientras que α_{ij} y β_i son elementos de un anillo K. El conjunto de estas condiciones se llama *sistema de ecuaciones lineales* con las incógnitas $\xi_1, \ldots, \xi_n$ sobre el anillo K. Una sucesión $\xi_1^0, \ldots, \xi_n^0$ de elementos del anillo K se llama *solución* del sistema (1) si al tomar en las condiciones (1) en lugar de las variables $\xi_1, \ldots, \xi_n$ los elementos correspondientes $\xi_1^0, \ldots, \xi_n^0$ todas las condiciones resultan verídicas. Introduciendo las matrices

$$A = \begin{bmatrix} \alpha_{11} & \ldots & \alpha_{1n} \\ \cdots & \cdots & \cdots \\ \alpha_{m1} & \ldots & \alpha_{mn} \end{bmatrix}, \quad x = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix} \quad \text{y} \quad b = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_m \end{bmatrix},$$

el sistema (1) puede ser representado en la siguiente forma matricial

$$Ax = b. \tag{2}$$

La matriz A se llama matriz del sistema (1) y la matriz

$$B = \begin{bmatrix} \alpha_{11} & \dots & \alpha_{1n} & \beta_1 \\ \dots & \dots & \dots & \dots \\ \alpha_{m1} & \dots & \alpha_{mn} & \beta_m \end{bmatrix}$$

se llama matriz *ampliada* del sistema (1). Las propiedades de la matriz B nos harán falta en el p. 5.3 al estudiar de forma más detallada las propiedades del sistema (1); por ahora nos limitaremos a considerar el caso en que $m = n$, es decir, en que la matriz A es cuadrada. Supongamos que la matriz A es invertible. En este caso, multiplicando ambos miembros de la igualdad (2) a la izquierda por A^{-1}, obtenemos

$$x = A^{-1}b. \tag{3}$$

Viceversa, multiplicando la relación (3) a la izquierda por A, obtenemos (2). Por consiguiente, las condiciones (2) y (3) son equivalentes y por ello la fórmula (3) puede ser considerada como la fórmula que ofrece la solución del sistema de ecuaciones (1) siempre que la matriz del sistema sea invertible.

Hasta el momento no hemos supuesto siquiera que el anillo K sea conmutativo. Supongamos ahora que K es un cuerpo conmutativo. En este caso, el hecho de que la matriz A sea invertible equivale a que sea regular; la matriz inversa se expresará en términos de la matriz adjunta en la forma anteriormente señalada y la fórmula (3) podrá ser representada de la forma siguiente

$$x = |A|^{-1}A^*b.$$

Tomando en lugar de la matriz A^* su expresión en términos de los adjuntos del determinante $|A|$ y realizando la multiplicación en el segundo miembro, obtenemos

$$\xi_i = \frac{|A|_{1i}\beta_1 + \dots + |A|_{ni}\beta_n}{|A|}$$

o bien, en forma definitiva,

$$\xi_i = \frac{1}{|A|} \begin{vmatrix} a_{11} & \dots & a_{1,\,i-1} & \beta_1 & \alpha_{1,\,i+1} & \dots & \alpha_{1n} \\ a_{21} & \dots & \alpha_{2,\,i-1} & \beta_2 & \alpha_{2,\,i+1} & \dots & \alpha_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \alpha_{n1} & \dots & \alpha_{n,\,i-1} & \beta_n & \alpha_{n,\,i+1} & \dots & \alpha_{nn} \end{vmatrix} (i = 1, \dots, n). \tag{4}$$

Para comprobar bastará desarrollar el determinante que figura en el numerador según los elementos de la i-ésima columna.

Un sistema de ecuaciones lineales sobre un cuerpo conmutativo se llama *cramériano* si el número de ecuaciones coincide con el

número de incógnitas y si el determinante de la matriz del sistema es diferente de cero. Vemos, por consiguiente, que *todo sistema crameriano de ecuaciones lineales tiene una solución, y sólo una,* $(\xi_1, \ldots, \xi_n)$ *y que esta solución viene dada por las fórmulas* (4). Las fórmulas (4) se llaman *fórmulas de Cramer*, en memoria de Cramer, matemático de mediados del siglo pasado, ya que aparecieron, por lo visto, por primera vez en uno de sus trabajos. Para $n = 2$ y $n = 3$ las fórmulas de Cramer han sido dadas de forma detallada en el p. 2.1.

Las fórmulas de Cramer dan una expresión «cerrada» de la solución de un sistema crameriano de ecuaciones en términos de los coeficientes de estas ecuaciones. Sin embargo, no es conveniente hallar la solución de un sistema lineal empleando las fórmulas de Cramer. En efecto, recurriendo a las fórmulas de Cramer, habrá que calcular los valores de $n+1$ determinantes de orden n. Según las estimaciones aproximadas, realizadas en el p. 2.2, habrá que efectuar para ello cerca de n^4 multiplicaciones y adiciones. Al mismo tiempo, el método corriente de eliminación sucesiva de incógnitas requiere, como ahora veremos, un número de operaciones considerablemente menor. Este método, ordenado de modo adecuado, se llama en la actualidad *algoritmo de Gauss*.

Consideremos, pues, de nuevo el sistema de ecuaciones lineales (1) cuyos coeficientes han sido tomados de un cuerpo conmutativo K. Si todos los coeficientes de una de las ecuaciones son iguales a cero podemos suprimirla del sistema. Si los coeficientes de las incógnitas son iguales a cero, mientras que $\beta \neq 0$, el sistema no tiene solución. Por esto se puede aceptar que algunos de los coeficientes de las incógnitas en la primera ecuación son diferentes de cero. En este caso, cambiando si es necesario la numeración de las incógnitas, podemos llevar el sistema a la forma (1) en la que $\alpha_{11} \neq 0$. Calculemos α_{11}^{-1} y multipliquemos todos los miembros de la primera ecuación por α_{11}^{-1}. A continuación, de la segunda, de la tercera, ..., de la m-ésima ecuaciones restamos término por término la primera ecuación multiplicada por $\alpha_{21}, \alpha_{31}, \ldots, \alpha_{m1}$, respectivamente. Después de esta operación nuestro sistema inicial resultará ser equivalente al sistema de tipo

$$\left.\begin{aligned} \xi_1 + \alpha'_{12}\xi_2 + \ldots + \alpha'_{1n}\xi_n &= \beta'_1, \\ \alpha'_{22}\xi_2 + \ldots + \alpha'_{2n}\xi_n &= \beta'_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots & \\ \alpha'_{m2}\xi_2 + \ldots + \alpha'_{mn}\xi_n &= \beta'_m. \end{aligned}\right\} \qquad (5)$$

Al pasar del sistema (1) al sistema (5) hemos realizado una inversión, nm multiplicaciones y $n(m-1)$ adiciones. Procediendo de la misma forma con las últimas $m-1$ ecuaciones del sistema (5), obtendremos — después de una inversión, de $(n-1)(m-1)$ multiplica-

ciones y de $(n-1)(m-2)$ adiciones—el sistema de tipo

$$\left.\begin{aligned} \xi_1 + \alpha'_{12}\xi_2 + \alpha'_{13}\xi_3 + \ldots + \alpha'_{1n}\xi_n &= \beta'_1, \\ \xi_2 + \alpha''_{23}\xi_3 + \ldots + \alpha''_{2n}\xi_n &= \beta''_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots & \\ \alpha^{(m-1)}_{m-1,3}\xi_3 + \ldots + \alpha^{(m-1)}_{m-1,n}\xi_n &= \beta^{(m-1)}_{m-1}. \end{aligned}\right\}$$

Continuando el proceso indicado, podremos encontrarnos con tres casos: a) en uno de los pasos llegamos a la conclusión de que el sistema no tiene solución; b) obtendremos un sistema de tipo

$$\left.\begin{aligned} \xi_1 + \gamma_{12}\xi_2 + \ldots + \gamma_{1r}\xi_r + \gamma_{1,r+1}\xi_{r+1} + \ldots + \gamma_{1n}\xi_n &= \delta_1, \\ \xi_2 + \ldots + \gamma_{2r}\xi_r + \gamma_{2,r+1}\xi_{r+1} + \ldots + \gamma_{2n}\xi_n &= \delta_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ \xi_r + \gamma_{r,r+1}\xi_{r+1} + \ldots + \gamma_{rn}\xi_n &= \delta_r; \end{aligned}\right\}$$

c) obtendremos un sistema de tipo

$$\left.\begin{aligned} \xi_1 + \gamma_{12}\xi_2 + \gamma_{13}\xi_3 + \ldots + \gamma_{1n}\xi_n &= \delta_1, \\ \xi_2 + \gamma_{23}\xi_3 + \ldots + \gamma_{2n}\xi_n &= \delta_2, \\ \ldots\ldots\ldots\ldots\ldots & \\ \xi_{n-1} + \gamma_{n-1,n}\xi_n &= \delta_{n-1}, \\ \xi_n &= \delta_n. \end{aligned}\right\} \qquad (6)$$

En el caso b), expresamos ξ_r, empleando la última ecuación, en términos de las incógnitas «libres» $\xi_{r+1}, \ldots, \xi_n$. Introduciendo el valor obtenido de ξ_r en la ecuación anterior, hallamos de ella ξ_{r-1}, etc. De esta forma encontraremos unas expresiones lineales de las incógnitas $\xi_1, \ldots, \xi_r$ en términos de las incógnitas «libres». Dando valores arbitrarios a las incógnitas «libres», podremos, a partir de las fórmulas señaladas, calcular los valores correspondientes de las incógnitas $\xi_1, \ldots, \xi_r$ y obtener de esta forma una solución $\xi_1, \ldots, \xi_r, \xi_{r+1}, \ldots, \xi_n$ del sistema dado. Lo mismo ocurrirá en el caso c); la diferencia consistirá solamente en que en este caso obtendremos unos valores determinados para todas las incógnitas $\xi_1, \ldots, \xi_n$.

Calculemos el número de operaciones que hay que efectuar en el caso c) para obtener la solución. Para reducir el sistema a la forma (6) (aceptando que $m = n$) tendremos que realizar n inversiones.

$$n^2 + (n-1)^2 + \ldots + 1^2 = \frac{1}{6}(2n+1)(n+1)n$$

multiplicaciones y

$$n(n-1) + (n-1)(n-2) + \ldots + 2 \cdot 1 = \frac{1}{3}(n+1)n(n-1)$$

adiciones. Además, para hallar del sistema triangular (6) las incógnitas habrá que realizar adicionalmente

$$1 + 2 + \ldots + (n-1) = \frac{1}{2}n(n-1)$$

multiplicaciones y un número igual de adiciones. Por consiguiente, para resolver el sistema (1) empleando el método de Gauss, es necesario realizar, en el caso general, n inversiones,

$$\frac{1}{3}n(n^2+3n-1) \approx \frac{1}{3}n^3$$

multiplicaciones y

$$\frac{1}{6}n(2n^2+3n-5) \approx \frac{1}{3}n^3$$

adiciones.

Observemos ahora que empleando el método de Gauss puede ser calculada también la matriz inversa A^{-1}. Esto se logra de la forma siguiente. Se escribe el sistema de ecuaciones lineales $Ax=b$, en el que las columnas x y b se consideran formadas por letras. Después, a partir del sistema dado y empleando el método de Gauss, las variables $\xi_1, \ldots, \xi_n$ se expresan en términos de las variables $\beta_1, \ldots, \beta_n$. Así se obtienen fórmulas de tipo

$$\xi_i = \gamma_{i1}\beta_1 + \ldots + \gamma_{in}\beta_n \quad (i = 1, \ldots, n).$$

Puesto que de $Ax=b$ se deduce que $x=A^{-1}b$, la matriz $\|\gamma_{ij}\|$ será precisamente la inversa de A. El cálculo del número de operaciones muestra que en este caso es suficiente realizar n inversiones, $\frac{1}{3}n(4n^2-1) \approx \frac{4}{3}n^3$ multiplicaciones y $\frac{2}{3}(n^2-n)(2n-1) \approx \frac{4}{3}n^3$ adiciones.

En los cálculos que hemos realizado acerca del número de operaciones necesarias para hallar el valor del determinante de una matriz, la solución de un sistema de ecuaciones lineales o la inversión de una matriz, no se han tenido en cuenta algunas operaciones secundarias, así como el aumento del número de cifras decimales en los números que se multiplican o se suman. Sin embargo, los resultados obtenidos pueden resultar útiles al decidir, si un sistema de ecuaciones lineales de interés práctico debe ser resuelto a mano o conviene pasar el encargo a un centro de cálculo.

Complementos y ejercicios

1. Empleando las fórmulas de Cramer, resuélvase el sistema

$$\left.\begin{aligned} x+y+z &= -2, \\ x+3y-2t &= -4, \\ 2x+z-t &= -1, \\ 2y-z-3t &= -3. \end{aligned}\right\}$$

2. Empleando el método de Gauss hállense las inversas de las matrices

$$\begin{bmatrix} 1 & 2 & -3 \\ 2 & -3 & 5 \\ 3 & -1 & 1 \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} a+1 & a & a & a \\ a & a+1 & a & a \\ a & a & a+1 & a \\ a & a & a & a+1 \end{bmatrix}.$$

3. Demuéstrese la fórmula (determinante de Vandermonde)

$$\begin{vmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \\ x_1^2 & x_2^2 & \dots & x_n^2 \\ \dots & \dots & \dots & \dots \\ x_1^{n-1} & x_2^{n-1} & \dots & x_n^{n-1} \end{vmatrix} = \prod_{i<j} (x_j - x_i).$$

4. Demuéstrese que

$$\begin{vmatrix} a+x_1 & a & \dots & a \\ a & a+x_2 & \dots & a \\ \dots & \dots & \dots & \dots \\ a & a & \dots & a+x_n \end{vmatrix} = x_1 x_2 \dots x_n \left(1 + \frac{a}{x_1} + \dots + \frac{a}{x_n}\right).$$

5. Si en el anillo principal K para cualquier α de $2\alpha = 0$ se deduce que $\alpha = 0$, el determinante de cualquier matriz antisimétrica sobre K de orden impar es igual a cero.

6. Consideremos una matriz cuadrada arbitraria A (sobre un anillo conmutativo K) de orden n. Suprimiendo en A las filas con los números $i_1, \dots, i_r$ ($1 \leqslant i_1 < \dots < i_r \leqslant n$) y las columnas con los números $j_1, \dots, j_r$ ($1 \leqslant j_1 < \dots < j_r \leqslant n$), obtendremos una matriz nueva $A^{i_1 \dots i_r}_{j_1 \dots j_r}$ de orden $n-r$. Los elementos de la matriz A que se hallan en los cruces de las filas y de las columnas suprimidas también forman una matriz cuadrada. El orden de esta última matriz es igual a r. Su determinante se denomina *menor* de orden r del determinante de la matriz A perteneciente a las filas con los números $i_1, \dots, i_r$ y a las columnas con los números $j_1, \dots, j_r$. Se llama *adjunto* de este menor el determinante de la matriz $A^{i_1 \dots i_r}_{j_1 \dots j_r}$ multiplicado por $(-1)^{s+t}$, donde $s = i_1 + \dots + i_r$ y $t = j_1 + \dots + j_r$. Tiene lugar la siguiente generalización del desarrollo de un determinante según los elementos de la primera y de la segunda filas: *si en un determinante de orden n se fijan r filas cualesquiera, la suma de los productos de los menores de orden r, pertenecientes a las filas escogidas, por sus adjuntos es igual al valor del determinante dado.*

7. Si la matriz cuadrada dada X es simétrica o antisimétrica, la matriz adjunta X^* también es simétrica o antisimétrica.

8. Sea A una matriz cuadrada de elementos enteros y sean x y b unas columnas, cuyos elementos son letras: $\xi_1, \dots, \xi_n$ y $\beta_1, \dots, \beta_n$, respectivamente. El sistema de ecuaciones lineales $Ax = b$ tiene soluciones enteras $\xi_1, \dots, \xi_n$ para cualesquiera valores enteros de $\beta_1, \dots, \beta_n$ si, y sólo si, $|A| = \pm 1$.

§ 3. Polinomios característico y mínimo

Todas las matrices que se consideran en este parrágrafo se suponen cuadradas y de un mismo orden n. Los elementos de estas matrices se toman de un cuerpo conmutativo K cualquiera, pero fijo.

3.1. Semejanza de matrices. La matriz A se llama *semejante* a la matriz B, si existe una matriz no degenerada X, tal, que

$$A = X^{-1}BX. \qquad (1)$$

En este caso también suele decirse que la matriz A es *la transformada de B por X*. Multiplicando la igualdad (1) a la izquierda por X y a la derecha por X^{-1}, obtenemos

$$B = XAX^{-1} = (X^{-1})^{-1} AX^{-1}.$$

Por consiguiente, *la semejanza de A con B implica la semejanza de B con A*. Además, si

$$A = X^{-1}BX \quad \text{y} \quad B = Y^{-1}CY,$$

realizando la sustitución obtenemos

$$A = (YX)^{-1}C(YX).$$

Por consiguiente, *dos matrices, semejantes cada una a otra tercera, son también semejantes*. Es obvio, finalmente, que toda matriz es semejante a sí misma.

Estas propiedades demuestran que el conjunto de todas las matrices cuadradas de un orden dado n, formadas por elementos de un cuerpo conmutativo K, se descompone de un modo natural en clases de matrices semejantes. Uno de los problemas centrales de la teoría de matrices es el de hallar las condiciones necesarias y suficientes de semejanza de matrices. La solución de este problema será dada en el capítulo IV. Aquí estableceremos solamente algunas propiedades previas de matrices semejantes.

Para transformar una suma de matrices por X es suficiente transformar por X cada sumando.

En efecto,

$$X^{-1}(A_1 + A_2 + \ldots + A_k)X = X^{-1}A_1X + X^{-1}A_2X + \ldots + X^{-1}A_kX.$$

Para transformar un producto de matrices por X es suficiente transformar cada factor

Efectivamente,

$$X^{-1}A_1X \cdot X^{-1}A_2X \ \ldots \ X^{-1}A_kX = X^{-1}A_1A_2 \ \ldots \ A_kX,$$

ya que los productos XX^{-1}, que aparecen en el primer miembro, son iguales a E y pueden ser omitidos.

Para transformar una potencia de una matriz es suficiente transformar la base de la potencia, es decir, se tiene

$$X^{-1}A^mX = (X^{-1}AX)^m.$$

Si $m \geqslant 0$, esta fórmula es un caso particular de la anterior. Si $m < 0$, tomemos $k = -m$. Entonces tendremos

$$X^{-1}A^mX = X^{-1}(A^{-1})^kX = (X^{-1}A^{-1}X)^k = (X^{-1}AX)^{-k} = (X^{-1}AX)^m.$$

El valor transformado de un polinomio en una matriz es igual al valor del polinomio en la matriz transformada, en otras palabras,

$$X^{-1}f(A)X = f(X^{-1}AX).$$

Esta proposición es resultado inmediato de las anteriores, ya que el valor de un polinomio en A se obtiene a partir de A realizando operaciones de elevación a potencias, de multiplicación por número y de adición.

Las reglas de la transformación de expresiones permiten en muchas ocasiones simplificar considerablemente los cálculos. Sean, por ejemplo,

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{y} \quad X = \begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix}.$$

Por inducción es fácil demostrar que

$$A^n = \begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix}.$$

Puesto que

$$X^{-1}AX = \begin{bmatrix} 7 & 4 \\ -9 & -5 \end{bmatrix},$$

tendremos, aplicando la regla de transformación de una potencia,

$$\begin{bmatrix} 7 & 4 \\ -9 & -5 \end{bmatrix}^n = X^{-1}A^nX = \begin{bmatrix} 1+6n & 4n \\ -9n & 1-6n \end{bmatrix}.$$

3.2. Polinomio característico. Sea A una matriz cuadrada y sean α_{ij} $(i, j = 1, 2, \ldots, n)$ sus elementos. La matriz

$$\lambda E - A = \begin{bmatrix} \lambda-\alpha_{11} & -\alpha_{12} & \ldots & -\alpha_{1n} \\ -\alpha_{21} & \lambda-\alpha_{22} & \ldots & -\alpha_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ -\alpha_{n1} & -\alpha_{n2} & \ldots & \lambda-\alpha_{nn} \end{bmatrix},$$

donde λ es una variable independiente, se llama *matriz característica* de A. Su determinante

$$\varphi(\lambda) = |\lambda E - A| \tag{1}$$

es evidentemente un polinomio en λ y se llama *polinomio característico* de la matriz A.

Para hallar los primeros términos de este polinomio recurramos a que el valor de un determinante es igual a la suma de los productos de sus elementos tomados de modo que haya uno de cada fila y uno de cada columna y provistos de signo adecuado. Para hallar el término de mayor grado respecto a λ es necesario tomar los productos de los elementos de mayor grado. Es nuestro caso este producto será uno solamente, a saber, el producto $(\lambda-\alpha_{11})(\lambda-\alpha_{22})\ldots(\lambda-\alpha_{nn})$ de los elementos diagonales. Todos los demás productos que componen el determinante serán de grado no mayor que $n-2$, ya que siendo $-\alpha_{ij}$ $(i \neq j)$ uno de los factores de alguno de estos productos, el último no contendrá los factores $\lambda-\alpha_{ii}$ y $\lambda-\alpha_{jj}$ y, por consiguiente, será de grado no mayor que $n-2$. De este modo, $\varphi(\lambda) = (\lambda-\alpha_{11})\ldots(\lambda-\alpha_{nn}) +$ términos de grado no mayor que $n-2$, o

$$\varphi(\lambda) = \lambda^n - (\alpha_{11} + \ldots + \alpha_{nn})\lambda^{n-1} + \ldots \tag{2}$$

La suma de los elementos diagonales de una matriz se llama *traza* de la matriz. Según la fórmula (2), *el grado del polinomio característico de una matriz es igual al orden de esta matriz, el coeficiente principal del polinomio característico es igual a 1, mientras que el coeficiente de λ^{n-1} es igual a la traza de la matriz tomada con el signo contrario.* Si en la fórmula (2) tomamos $\lambda = 0$, tendremos

$$\varphi(0) = |-A| = (-1)^n |A|.$$

Pero $\varphi(0)$ es el término independiente del polinomio característico. Es decir, el término independiente del polinomio característico de una matriz A es igual al determinante de esta matriz multiplicado por $(-1)^n$, donde n es el orden de la matriz A.

El teorema que sigue describe una de las propiedades más importantes del polinomio característico.

TEOREMA 1. *Los polinomios característicos de matrices semejantes coinciden.*

En efecto, sea A una matriz semejante a la matriz B:

$$A = X^{-1}BX.$$

Entonces para el polinomio característico de A tenemos

$$|\lambda E - A| = |\lambda E - X^{-1}BX| = |X^{-1}(\lambda E - B)X| = |X^{-1}| \cdot |\lambda E - B| \cdot |X|.$$

Los determinantes $|X^{-1}|$ y $|X|$ son números recíprocos cuyo producto es igual a la 1; por esto, se tiene

$$|\lambda E - B| = |\lambda E - A|$$

que es lo que se quería demostrar.

De este teorema se deduce, en particular, que *las matrices semejantes tienen trazas y determinantes iguales*, ya que la traza y el determinante de una matriz, provistos de signo adecuado, son coeficientes de su polinomio característico.

La igualdad de los polinomios característicos es una condición necesaria pero, como regla general, no suficiente de la semejanza de matrices. Por ejemplo, los polinomios característicos de las matrices

$$E = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{y} \quad A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

coinciden. Sin embargo, A no puede ser semejante a E ya que para cualquier matriz X se tiene

$$X^{-1}EX = X^{-1}X = E.$$

Las raíces del polinomio característico de una matriz se llaman *números característicos* o *valores propios* de la misma. Las raíces múltiples del polinomio característico se llaman valores propios múltiples de la matriz. Es conocido que la suma de todas las raíces

reales y complejas de un polinomio de grado n, cuyo coeficiente principal es la 1, es igual al coeficiente de la $(n-1)$-ésima potencia de la variable tomado con el signo contrario. De la fórmula (2) resulta, por ello, que *en el cuerpo de los números complejos la suma de todos los valores propios de una matriz es igual a su traza.*

Observemos que la traza de una suma de matrices es igual a la suma de las trazas de los sumandos y que la traza del producto de un número por una matriz es igual al producto de este número por la traza de la matriz. Ambas afirmaciones pueden ser resumidas en una fórmula:

$$\text{traza}\,(\alpha A+\beta B)=\alpha\cdot\text{traza}\,A+\beta\cdot\text{traza}\,B$$

para cuya demostración basta considerar las matrices correspondientes y calcular sus trazas.

En el p. 1.2 a todo polinomio $\varphi(\lambda)$ ha sido puesta en correspondencia una matriz $\varphi(A)$ llamada valor del polinomio $\varphi(\lambda)$ para $\lambda=A$. Si $\varphi(A)=O$, se dice que A es *raíz* de $\varphi(\lambda)$.

TEOREMA DE HAMILTON-CAYLEY. *Toda matriz es raíz de su polinomio característico.*

Sea A una matriz. Sea B la matriz adjunta de la matriz característica $\lambda E-A$ (véase el p. 2.3). Sean β_{ij} $(i, j=1, 2, \ldots, n)$ los elementos de la matriz B. Estos elementos son los adjuntos del determinante $|\lambda E-A|$ y, por lo tanto, representan polinomios en λ de grado no superior a $n-1$. Sea

$$\beta_{ij}=\beta_{ij}^{(0)}+\beta_{ij}^{(1)}\lambda+\ldots+\beta_{ij}^{(n-1)}\lambda^{(n-1)}.$$

Consideremos las matrices numéricas auxiliares

$$B^{(k)}=\begin{bmatrix}\beta_{11}^{(k)} & \beta_{12}^{(k)} & \ldots & \beta_{1n}^{(k)}\\ \beta_{21}^{(k)} & \beta_{22}^{(k)} & \ldots & \beta_{2n}^{(k)}\\ \ldots & \ldots & \ldots & \ldots\\ \beta_{n1}^{(k)} & \beta_{n2}^{(k)} & \ldots & \beta_{nn}^{(k)}\end{bmatrix}\quad (k=0,\ 1,\ \ldots,\ n-1).$$

Es evidente que la matriz B puede ser representada en este caso en la forma siguiente:

$$B=B^{(0)}+\lambda B^{(1)}+\ldots+\lambda^{n-1}B^{(n-1)}.$$

En virtud de la propiedad principal de las matrices adjuntas, se tiene

$$B(\lambda E-A)=|\lambda E-A|\cdot E. \tag{3}$$

Aquí $|\lambda E-A|$ es el polinomio característico de la matriz A que representaremos mediante $\varphi(\lambda)$. Sea

$$\varphi(\lambda)=\alpha_0+\alpha_1\lambda+\ldots+\alpha_{n-1}\lambda^{n-1}+\lambda^n.$$

Ahora podemos escribir la igualdad (3) en forma más detallada:

$$(B^{(0)}+\lambda B^{(1)}+\ldots+\lambda^{n-1}B^{(n-1)})(\lambda E-A)=$$
$$=(\alpha_0+\alpha_1\lambda+\ldots+\alpha_{n-1}\lambda^{n-1}+\lambda^n)\cdot E.$$

Suprimiendo los paréntesis y comparando los coeficientes de potencias iguales de λ, obtenemos

$$\begin{aligned}
-B^{(0)}A &= \alpha_0 E,\\
-B^{(1)}A+B^{(0)} &= \alpha_1 E,\\
-B^{(2)}A+B^{(1)} &= \alpha_2 E,\\
&\cdots\cdots\cdots\cdots\\
-B^{(n-1)}A+B^{(n-2)} &= \alpha_{n-1}E,\\
B^{(n-1)} &= E.
\end{aligned}$$

Multipliquemos a la derecha estas igualdades por E, A, ..., A^n, respectivamente, y sumémoslas. Todos los términos del primer miembro se anularán y obtendremos

$$O=\alpha_0 E+\alpha_1 A+\alpha_2 A^2+\ldots+A^n,$$

es decir, $\varphi(A)=O$ que es lo que se quería.

3.3. Polinomio mínimo. Consideremos todos los polinomios no nulos $f(\lambda)$ para los cuales una matriz A dada es una raíz. Existen polinomios de este tipo; entre ellos figura, por ejemplo, el polinomio característico de la matriz A. El polinomio no nulo de menor grado y de coeficiente principal igual a la 1, para el cual la matriz A es una raíz, se llama *polinomio mínimo* de esta matriz.

Toda matriz A tiene sólo un polinomio mínimo. En efecto, si hubiese dos, digamos $\psi_1(\lambda)$ y $\psi_2(\lambda)$, la diferencia $\psi_1(\lambda)-\psi_2(\lambda)$ sería un polinomio no nulo de grado menor, para el cual la matriz A también sería una raíz. Dividiendo esta diferencia por su coeficiente principal, tendríamos un polinomio de coeficiente principal igual a la 1, la matriz A sería una de sus raíces y su grado sería inferior al de los polinomios mínimos $\psi_1(\lambda)$ y $\psi_2(\lambda)$, lo cual estaría en contradicción con la definición de los polinomios mínimos.

Todo polinomio $f(\lambda)$ para el cual la matriz A es una raíz es divisible por el polinomio mínimo $\psi(\lambda)$ de esta matriz.

Efectivamente, supongamos, al contrario, que $f(\lambda)$ no es divisible por $\psi(\lambda)$. Designando mediante $q(\lambda)$ el cociente y mediante $r(\lambda)$ el resto de la división de $f(\lambda)$ por $\psi(\lambda)$, tendremos

$$f(\lambda)=\psi(\lambda)\,q(\lambda)+r(\lambda).$$

Tomando aquí $\lambda=A$ y valiéndonos de que $\psi(A)=f(A)=O$, obtendremos $r(A)=O$. Pero el grado del resto $r(\lambda)$ es menor que el grado del divisor $\psi(\lambda)$. Por consiguiente, $r(\lambda)$ es un polinomio no nulo, la matriz A es su raíz y su grado es inferior al grado del polinomio mínimo $\psi(\lambda)$, lo cual es contradictorio. Hemos

demostrado nuestra afirmación. En particular, *el polinomio mínimo de una matriz es un divisor de su polinomio característico.*

Sabemos que las matrices semejantes tienen un mismo polinomio característico. Esta misma propiedad la tiene el polinomio mínimo: las matrices semejantes tienen polinomios mínimos idénticos. En efecto, sea A semejente a B: $A = X^{-1}BX$. Si $f(\lambda)$ es un polinomio que tiene a B como raíz, tendremos en virtud del p. 3.1

$$f(A) = f(X^{-1}BX) = X^{-1}f(B)X = O.$$

Por consiguiente, el conjunto de polinomios que tienen como raíz una de las matrices semejantes coincide con el conjunto de polinomios que tienen como raíz otra de las matrices semejantes. Por esto, el polinomio de grado menor y de coeficiente principal igual a la 1, perteneciente a este conjunto, será el polinomio mínimo de ambas matrices.

La igualdad de los polinomios mínimos es una condición necesaria más de la semejanza de matrices. Sin embargo, esta condición tampoco es suficiente. Consideremos por ejemplo, las matrices

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix} \quad \text{y} \quad B = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

Sus polinomios característicos son iguales, respectivamente, a

$$(\lambda - 2)(\lambda - 3)^2 \quad \text{y} \quad (\lambda - 2)^2(\lambda - 3).$$

Puesto que estos polinomios son diferentes, las matrices A y B no son semejantes. El polinomio mínimo de la matriz A debe ser divisor de su polinomio característico, es decir, debe coincidir con uno de los polinomios siguientes: $\lambda - 2$, $\lambda - 3$, $(\lambda - 3)^2$, $(\lambda - 2)(\lambda - 3)$ o $(\lambda - 2)(\lambda - 3)^2$. Tomando aquí en lugar de λ la matriz A, encontraremos que el cero se obtiene por primera vez en el caso del polinomio $(\lambda - 2)(\lambda - 3)$. Por consiguiente, este polinomio será precisamente el polinomio mínimo de la matriz A. De la misma forma determinaremos que el polinomio mínimo de la matriz B es también el polinomio $(\lambda - 2)(\lambda - 3)$. Es decir, los polinomios mínimos de las matrices A y B coinciden, mientras que las matrices A y B no son semejantes.

En el p. 1.4 hemos dicho que las matrices de tipo

$$A = \begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_s \end{bmatrix},$$

donde $A_1, \ldots, A_s$ son matrices cuadradas, se llaman *descompuestas en células* $A_1, \ldots, A_s$ o *descompuestas en la suma directa de estas células.*

El polinomio característico de una matriz descompuesta es igual al producto de los polinomios característicos de sus células diagonales.

En efecto, si A se descompone en células $A_1, \ldots, A_s$, es fácil ver que su matriz característica es de la forma

$$\lambda E - A = \begin{bmatrix} \lambda E_1 - A_1 & & & \\ & \lambda E_2 - A_2 & & \\ & & \ddots & \\ & & & \lambda E_s - A_s \end{bmatrix},$$

donde $E_1, \ldots, E_s$ son las matrices unidades de ordenes correspondientes. Se sabe del p. 2.1 que el determinante de una matriz descompuesta es igual al producto de los determinantes de sus células diagonales. Por consiguiente,

$$|\lambda E - A| = |\lambda E_1 - A_1| \cdot |\lambda E_2 - A_2| \ldots |\lambda E_s - A_s|$$

que es lo que se quería demostrar.

El polinomio mínimo de una matriz descompuesta es igual al mínimo común múltiplo de los polinomios mínimos de sus células diagonales.

Supongamos que la matriz A se descompone en las células $A_1, \ldots, A_s$. Sean $\psi_1(\lambda), \ldots, \psi_s(\lambda)$ sus polinomios mínimos respectivos. Consideremos un polinomio arbitrario $f(\lambda)$. Si $f(A) = O$, de la fórmula (7) del p. 1.4 resulta que $f(A_1) = \ldots = f(A_s) = O$. Pero todo polinomio que tiene como raíz a la matriz A_i es divisible por el polinomio mínimo $\psi_i(\lambda)$ de esta matriz. Por consiguiente, $f(\lambda)$ es un común múltiplo de los polinomios $\psi_1(\lambda), \ldots, \psi_s(\lambda)$. Viceversa, si un polinomio cualquiera $f(\lambda)$ es un común múltiplo de $\psi_1(\lambda), \ldots, \psi_s(\lambda)$, se tiene evidentemente $f(A) = O$. Por consiguiente, para obtener el polinomio mínimo de la matriz A es preciso tomar el común múltiplo de menor grado de los polinomios $\psi_1(\lambda), \ldots, \psi_s(\lambda)$, es decir, el mínimo común múltiplo, que es lo que se quería demostrar.

Una matriz celular A se llama *semidescompuesta* o *celular triangular*, si todas sus células diagonales son cuadradas, mientras que las células que figuran a un lado cualquiera de la diagonal principal están formadas por ceros. En lo sucesivo siempre aceptaremos que las células nulas de la matriz semidescompuesta se hallan sobre la diagonal principal. Por consiguiente, las matrices celulares

semidescompuestas tienen la estructura siguiente

$$A = \begin{bmatrix} A_{11} & & & \\ A_{21} & A_{22} & & \\ \cdots & \cdots & \cdots & \cdots \\ A_{s1} & A_{s2} & \cdots & A_{ss} \end{bmatrix}, \tag{1}$$

donde $A_{11}, \ldots, A_{ss}$ son células cuadradas y todas las células que se hallan encima de éstas están formadas por ceros.

Sea B otra matriz semidescompuesta cualquiera, cuyas células diagonales son del mismo orden que las células correspondientes de la matriz A. En virtud de las reglas de operación con matrices celulares, tenemos

$$\begin{bmatrix} A_{11} & & & \\ A_{21} & A_{22} & & \\ \cdots & \cdots & \cdots & \cdots \\ A_{s1} & A_{s2} & \cdots & A_{ss} \end{bmatrix} + \begin{bmatrix} B_{11} & & & \\ B_{21} & B_{22} & & \\ \cdots & \cdots & \cdots & \cdots \\ B_{s1} & B_{s2} & \cdots & B_{ss} \end{bmatrix} =$$

$$= \begin{bmatrix} A_{11}+B_{11} & & & \\ A_{21}+B_{21} & A_{22}+B_{22} & & \\ \cdots & \cdots & \cdots & \cdots \\ A_{s1}+B_{s1} & A_{s2}+B_{s2} & \cdots & A_{ss}+B_{ss} \end{bmatrix},$$

$$\begin{bmatrix} A_{11} & & & \\ A_{21} & A_{22} & & \\ \cdots & \cdots & \cdots & \cdots \\ A_{s1} & A_{s2} & \cdots & A_{ss} \end{bmatrix} \cdot \begin{bmatrix} B_{11} & & & \\ B_{21} & B_{22} & & \\ \cdots & \cdots & \cdots & \cdots \\ B_{s1} & B_{s2} & \cdots & B_{ss} \end{bmatrix} = \begin{bmatrix} A_{11}B_{11} & & & \\ C_{21} & A_{22}B_{22} & & \\ \cdots & \cdots & \cdots & \cdots \\ C_{s1} & C_{s2} & \cdots & A_{ss}B_{ss} \end{bmatrix},$$

donde $C_{ij} = A_{ij}B_{jj} + A_{i,j+1}B_{j+1,j} + \ldots + A_{ii}B_{ij}$. Por consiguiente, *la suma y el producto de matrices semidescompuestas son matrices semidescompuestas cuyas células diagonales son iguales a las sumas y a los productos de las células correspondientes de las matrices dadas.* En particualr, si $f(\lambda)$ es un polinomio en λ y A es una matriz semidescompuesta de tipo (1), se tiene

$$f(A) = \begin{bmatrix} f(A_{11}) & & & \\ D_{21} & f(A_{22}) & & \\ \cdots & \cdots & \cdots & \cdots \\ D_{s1} & D_{s2} & \cdots & f(A_{ss}) \end{bmatrix} \tag{2}$$

(las células D_{ij} tienen una estructura más compleja). Del § 2 conocemos que el determinante de una matriz semidescompuesta es igual al producto de los determinantes de sus células diagonales.

De aquí se deduce directamente que *el polinomio característico de una matriz semidescompuesta es igual al producto de los polinomios característicos de las células diagonales de esta matriz.*

Si las células de una matriz semidescompuesta A son de orden 1, se dice que A tiene *forma triangular* o, simplemente, que es *una matriz triangular*. Los valores propios de una matriz triangular son

iguales a sus elementos diagonales. De las fórmulas (2) resulta también que siendo $\zeta_1, \zeta_2, \ldots, \zeta_n$ los valores propios de una matriz triangular, los valores propios de la matriz $f(A)$ serán $f(\zeta_1)$, $f(\zeta_2), \ldots, f(\zeta_n)$.

Ejemplos y problemas

1. Una matriz se llama *nilpotente* si una de sus potencias es igual a la matriz nula. Demuéstrese que una matriz semidescompuesta es nilpotente si, y sólo si, son nilpotentes sus células diagonales.

2. Empleando el teorema de Hamilton-Cayley demuéstrese que si en el cuerpo de los números complejos todos los valores propios de una matriz son iguales a cero, la matriz es nilpotente.

3. Demuéstrese que si en la suma *directa* de matrices se cambian entre sí los sumandos, la suma nueva será una matriz *semejante* a la suma inicial.

4. Calcúlense los polinomios característicos y los valores propios de las matrices

$$A = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} \cos\alpha & \operatorname{sen}\alpha \\ -\operatorname{sen}\alpha & \cos\alpha \end{bmatrix} \quad \text{y} \quad C = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \\ 3 & 3 & 6 \end{bmatrix}.$$

5. Hállense los polinomios mínimos de las matrices

$$\begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}, \quad \begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \quad \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}.$$

6. Demuéstrese que los valores propios de una matriz diagonal son sus elementos diagonales.

7. Demuéstrese la fórmula

$$\text{traza}\,(AB) = \text{traza}\,(BA).$$

8. Si en una matriz cuadrada de orden n se suprimen m filas con los números $i_1, i_2, \ldots, i_m$ y m columnas con los mismos números, quedará una matriz de orden $n-m$ y su determinante se llama *menor principal* de orden $n-m$. Demuéstrese que el coeficiente de λ^m en el polinomio característico de la matriz A es igual a la suma de sus menores principales de orden $n-m$ multiplicada por $(-1)^{n-m}$.

9. Demuéstrese que el polinomio característico de la matriz AB coincide con el polinomio característico de la matriz BA.

10. Si $\zeta_1, \zeta_2, \ldots, \zeta_n$ son los valores propios de una matriz A, los valores propios de la matriz $f(A)$, donde $f(\lambda)$ es un polinomio, serán iguales a $f(\zeta_1)$, $f(\zeta_2), \ldots, f(\zeta_n)$.

Capítulo II

Espacios lineales

§ 4. Dimensión

4.1. Módulos y espacios vectoriales. Un conjunto arbitrario no vacío de elementos $\mathfrak{L}$ se llama *módulo sobre un anillo* K, si se cumplen las condiciones siguientes:

a) existe una regla que a partir de cualquier par de elementos a y b de $\mathfrak{L}$ permite hallar un elemento de $\mathfrak{L}$ que se llama *suma* de los dos primeros y se designa mediante $a+b$;

b) existe una regla que a partir de cualquier número α de K y de cualquier elemento a de $\mathfrak{L}$ permite hallar en $\mathfrak{L}$ un elemento nuevo que se llama *producto de* α *por* a y que se designa mediante αa;

c) las operaciones de adición y de multiplicación por número satisfacen los axiomas siguientes:

1° la adición es conmutativa:

$$a+b=b+a;$$

2° la adición es asociativa:

$$a+(b+c)=(a+b)+c;$$

3° es posible realizar la sustracción, es decir, para todo par de elementos a y b de $\mathfrak{L}$ existe en $\mathfrak{L}$ un elemento x tal que

$$a+x=b;$$

4° la multiplicación es asociativa:

$$\alpha(\beta a)=(\alpha\beta)a;$$

5° la multiplicación es distributiva respecto a la adición en $\mathfrak{L}$:

$$\alpha(a+b)=\alpha a+\alpha b;$$

6° la multiplicación es distributiva respecto a la adición de números:

$$(\alpha+\beta)a=\alpha a+\beta a.$$

Si el anillo K está provisto del elemento unidad 1, suele exigirse que se cumpla además la condición

7° $\qquad 1 \cdot a = a \quad (a \in \mathfrak{L}).$

En este caso el módulo se llama *unitario*. Un módulo unitario sobre un cuerpo K se llama *espacio lineal* o *vectorial* sobre K. Los elementos de los módulos y de los espacios vectoriales se llaman *vectores* y en lo que sigue se representan con letras latinas minúsculas a, b, x, y, ... Los espacios lineales y los conjuntos de vectores se representarán con letras góticas mayúsculas $\mathfrak{A}$, $\mathfrak{B}$, $\mathfrak{L}$, $\mathfrak{M}$, ...

En la definición dada más arriba los vectores se multiplican a la izquierda por los elementos del anillo K. Por esta razón, los módulos y espacios definidos más arriba se llaman también módulos a la izquierda y espacios a la izquierda sobre K. Si se exige que estén definidos los productos a la derecha de los elementos de $\mathfrak{L}$ por los elementos de K y si se modifican de modo correspondiente los axiomas de 4° a 7°, se obtendrá una estructura que se llama módulo a la derecha y, respectivamente, espacio lineal a la derecha. Está claro que las propiedades de los espacios a la izquierda y a la derecha son las mismas; tiene importancia distinguir la multiplicación a la izquierda y a la derecha sólo cuando estén definidas simultáneamente. En lo que sigue los espacios a la izquierda sobre K se llaman simplemente espacios sobre K.

Un espacio lineal sobre el cuerpo de los números complejos se llama *espacio lineal complejo* y un espacio lineal sobre el cuerpo de los números reales se llama *espacio lineal real*. Los ejemplos principales de módulos y de espacios lineales se darán más tarde, mientras que ahora consideraremos los corolarios más simples que se desprenden inmediatamente de los axiomas de 1° a 7°.

Ante todo, la condición a) y los axiomas 2° y 3° muestran que todo módulo es un grupo respecto a la operación de adición de vectores y, además, según el axioma 1° este grupo debe ser conmutativo. Por consiguiente, igual que en cualquier grupo conmutativo, la suma de un número finito de vectores no depende ni del orden de los sumandos de esta suma ni de la forma en la que están dispuestos los paréntesis. Por ejemplo,

$$(a+b)+(c+d) = c+(a+(d+b)).$$

Además, entre los vectores de un módulo arbitrario existe, al igual que en todo grupo aditivo, un único vector—designémoslo mediante o—que posee la propiedad de que

$$x+o=o+x=x$$

cualquiera que sea el vector x del módulo considerado. El vector o se llama *vector nulo* o simplemente *cero* del módulo. Además, para todo vector a del módulo $\mathfrak{L}$ existe en el último un vector y,

y sólo uno, que satisface la ecuación

$$a+y=o.$$

Este vector y se indica por $-a$ y se llama *opuesto* de a.

De los axiomas 1°, 2° y 3° se desprende que $-o=o$ y que para cualesquiera a y b

$$-(-a)=a \quad \text{y} \quad -(a+b)=(-a)+(-b).$$

En las expresiones de tipo $(-a)+b+(-c)$ suelen omitirse los paréntesis escribiendo $-a+b-c$. La operación binaria $-$, definida mediante la fórmula

$$x-y=x+(-y),$$

se llama operación de *sustracción* de vectores. El hecho de que un mismo símbolo indica dos operaciones diferentes (la del paso al vector opuesto y la operación binaria de sustracción) no originará en lo sucesivo ningún inconveniente.

Hasta el momento hemos considerado sólo aquellas propiedades de los módulos que los caracterizan por ser ellos grupos conmutativos respecto a la adición, es decir, aquellas propiedades que se desprenden de los axiomas 1°, 2° y 3°. Teniendo en cuenta los axiomas de 4° a 7°, obtenemos fácilmente las identidades

$$0\cdot a=o,$$

$$ma=a+a+\ldots+a$$

(m sumandos; m es un número entero positivo),

$$(-\alpha)\cdot a=-(\alpha a),$$

$$\alpha\cdot o=o.$$

En efecto, la primera es consecuencia de que

$$a=(1+0)a=1\cdot a+0\cdot a=a+0\cdot a;$$

la segunda se demuestra de modo siguiente:

$$ma=(1+\ldots+1)a=1\cdot a+\ldots+1\cdot a=a+\ldots+a.$$

Luego, puesto que

$$\alpha a+(-\alpha)a=(\alpha-\alpha)a=o,$$

se tiene $(-\alpha)a=-(\alpha a)$. Finalmente, la cuarta igualdad es consecuencia de que

$$\alpha\cdot o+\alpha a=\alpha(o+a)=\alpha a.$$

Observemos también que, siendo K un cuerpo, de $\alpha a=o$ se deduce que o bien $\alpha=0$ o bien $a=o$. En efecto, si $\alpha\neq 0$, multiplicando la relación $\alpha a=o$ por α^{-1} obtenemos $a=o$.

Las expresiones de tipo

$$\alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_s a_s$$

se llaman *combinaciones lineales* de los vectores $a_1, a_2, \ldots, a_s$. De nuestras consideraciones se desprende que las combinaciones lineales pueden ser sumadas, restadas y multiplicadas por números y que en ellas se pueden reducir los términos semejantes siguiendo las reglas habituales.

El ejemplo más conocido de un espacio lineal es el conjunto de los segmentos orientados que parten de un punto fijo O de nuestro espacio habitual. Multiplicar un segmento por un número real positivo α significa aumentar su longitud α veces sin alterar su dirección. Si α es negativo, la multiplicación de un segmento por α significa que su longitud aumenta $|\alpha|$ veces y que su dirección cambia por la contraria. Análogamente, sumar dos segmentos significa tomar la diagonal del paralelogramo construido a partir de estos segmentos. El vector nulo será el segmento, cuyo origen y extremo coinciden con el punto O. Puesto que la operación de multiplicación de un segmento por un número está definida sólo para los números reales, el campo principal K es en este caso el cuerpo de todos los números reales y el espacio de segmentos orientados es un espacio lineal *real*. Lo llamaremos siempre espacio de vectores-segmentos corrientes y lo indicaremos por $\mathfrak{R}$.

Un ejemplo más general de espacios lineales, y además básico para toda la teoría de los mismos, es *el espacio de filas*. Consideremos el conjunto de todas las sucesiones de tipo $[\alpha_1, \alpha_2, \ldots, \alpha_n]$, donde $\alpha_1, \ldots, \alpha_n$ son números de un cuerpo K y n es un número entero dado. También llamaremos estas sucesiones *filas*, considerándolas como matrices compuestas de una fila. Dos filas se llaman *iguales* si son iguales sus elementos respectivos. Las operaciones de adición de filas y de multiplicación de filas por un número se definen mediante las fórmulas matriciales correspondientes:

$$\beta[\alpha_1, \ldots, \alpha_n] = [\beta\alpha_1, \ldots, \beta\alpha_n],$$
$$[\alpha_1, \alpha_2, \ldots, \alpha_n] + [\beta_1, \beta_2, \ldots, \beta_n] =$$
$$= [\alpha_1 + \beta_1, \alpha_2 + \beta_2, \ldots, \alpha_n + \beta_n].$$

Está claro que en este caso los axiomas de 1° a 7° se cumplen y que el conjunto de todas las filas de longitud n formadas por elementos de un cuerpo K es *un espacio lineal sobre* K. Siendo K un anillo cualquiera, obtenemos un ejemplo de *un módulo sobre* K.

En lugar de filas se pueden considerar matrices de un número cualquiera, pero fijo, de filas y de columnas con elementos de un anillo K. De acuerdo con las reglas del cálculo de matrices, cualquier matriz de este tipo puede ser multiplicada por un número de K y cualesquiera dos pueden ser sumadas, obteniéndose ambas veces una matriz del mismo tipo. Es obvio que los axiomas de 1° a

7° también se cumplen aquí y, por consiguiente, las matrices de m filas y de n columnas constituyen, respecto a las operaciones de adición y de multiplicación por número, un módulo sobre K o un espacio lineal, si K es un cuerpo.

Consideremos un ejemplo más. Sea $\mathfrak{M}$ un conjunto arbitrario de cualesquiera elementos y sea K un anillo. Supongamos que tenemos a nuestra disposición una ley que pone en correspondencia a cada elemento m de $\mathfrak{M}$ un número determinado de K. Toda ley de esta índole se llama *función* definida sobre el conjunto $\mathfrak{M}$ y con valores en K. Si la función se indica con una letra cualquiera f, mediante $f(m)$ suele indicarse el número que corresponde al elemento m. El número $f(m)$ se llama valor de la función f sobre el elemento m. Dos funciones f y g que satisfacen la igualdad $f(m) = g(m)$ cualquiera que sea m de $\mathfrak{M}$, se llaman *iguales*. Para las funciones definidas sobre $\mathfrak{M}$ se pueden introducir de modo habitual las operaciones de multiplicación por número, de adición y de multiplicación de funciones. Por ejemplo, dados un número α y una función f y poniendo en correspondencia a cada elemento m de $\mathfrak{M}$ el número $\alpha f(m)$, obtenemos una función nueva que se llama producto del número α por la función f. Análogamente se definen la suma y el producto de dos funciones. Si consideramos sólo las dos primeras de estas operaciones — la multiplicación por número y la adición — es fácil ver que se cumplen los axiomas de un módulo. Por consiguiente, el conjunto de todas las funciones, definidas sobre un conjunto dado $\mathfrak{M}$ y con valores en un anillo K, constituye *un módulo sobre* K. Siendo el anillo inicial un cuerpo, obtenemos, al igual que anteriormente, un ejemplo de *espacio vectorial*.

En el caso en que $\mathfrak{M}$ consta de un número finito de elementos, se puede introducir una notación cómoda para las funciones definidas sobre $\mathfrak{M}$. Sean $m_1, m_2, \ldots, m_s$ los elementos del conjunto $\mathfrak{M}$. Indiquemos mediante $[m_i]$ la función que es igual a 1, sobre m_i e igual a 0 sobre todos los demás elementos de $\mathfrak{M}$ ($i = 1, \ldots, \ldots, s$). En estas condiciones el producto $\alpha[m_i]$ será la función igual a α sobre m_i e igual a 0 sobre todos los demás elementos de $\mathfrak{M}$ y la expresión

$$\alpha_1[m_1] + \alpha_2[m_2] + \ldots + \alpha_s[m_s] \qquad (1)$$

será, obviamente, la función que es igual a α_1 sobre m_1, a α_2 sobre m_2, ..., a α_s sobre m_s. Por consiguiente, toda función definida sobre $\mathfrak{M}$ puede ser representada en la forma (1) y esta representación, como es fácil ver, es única. Si no existe el peligro de confusión, los paréntesis en (1) suelen omitirse y en lugar de (1) suele escribirse brevemente:

$$\alpha_1 m_1 + \alpha_2 m_2 + \ldots + \alpha_s m_s$$

con la particularidad de que los términos de coeficientes nulos no se escriben. Por ejemplo, si $\mathfrak{M}$ se compone de las letras a, b y c, la expresión $2a-c$ significa la función que es igual a 2 sobre a, a cero sobre b y a -1 sobre c.

4.2. Dependencia lineal. Sea $\mathfrak{L}$ un espacio vectorial sobre un cuerpo de coeficientes K y sean $a_1, a_2, \ldots, a_m$ unos vectores de este espacio. La relación de tipo

$$\alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_m a_m = o,$$

donde $\alpha_1, \ldots, \alpha_m$ son números de K, se llama *relación de dependencia lineal* entre los vectores $a_1, \ldots, a_m$. Si todos los coeficientes $\alpha_1, \ldots, \alpha_m$ son iguales a cero, la relación se llama *trivial*. En el caso contrario, es decir, si al menos uno de los coeficientes es diferente de cero, la relación se llama *no trivial*. Está claro que una relación trivial existe entre cualesquiera vectores. La respuesta a la pregunta sobre si existe o no una relación no trivial, depende de los vectores que se consideran. Por ejemplo, en el espacio de filas de longitud tres entre los vectores $a=[1, 4, 6]$, $b=[1, -1, 1]$ y $c=[1, 1, 3]$ existe la siguiente relación de dependencia lineal

$$2a+3b-5c=o.$$

Por otra parte, en este mismo espacio no existe ninguna relación no trivial entre los vectores $e_1=[1, 0, 0]$, $e_2=[0, 1, 0]$ y $e_3=[0, 0, 1]$ ya que la relación

$$\alpha_1 e_1 + \alpha_2 e_2 + \alpha_3 e_3 = o$$

significa que

$$[\alpha_1, \alpha_2, \alpha_3] = [0, 0, 0],$$

de donde se tiene

$$\alpha_1 = \alpha_2 = \alpha_3 = 0.$$

Un sistema finito de vectores $a_1, a_2, \ldots, a_m$ de un espacio lineal se llama linealmente dependiente, si entre ellos existe una relación no trivial.

Si no existe tal relación, es decir, si de toda relación de tipo

$$\alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_m a_m = o$$

se deduce que $\alpha_1 = \alpha_2 = \ldots = \alpha_m = 0$, el sistema $a_1, a_2, \ldots, a_m$ se llama *linealmente independiente*.

De esta definición se deduce inmediatamente que *si a un sistema linealmente dependiente de vectores $a_1, a_2, \ldots, a_m$ se agregan otros vectores $b_1, \ldots, b_k$, el sistema ampliado seguirá siendo linealmente dependiente*. En efecto, si

$$\alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_m a_m = o$$

es una relación no trivial entre $a_1, \ldots, a_m$, tendremos que

$$\alpha_1 a_1 + \ldots + \alpha_m a_m + 0 \cdot b_1 + \ldots + 0 \cdot b_k = o$$

será una relación no trivial entre $a_1, \ldots, a_m, b_1, \ldots, b_k$.

En los problemas relacionados con la dependencia lineal el vector nulo o ocupa una posición especial debido a que todo sistema que contenga el vector nulo es linealmente dependiente. Para demostrar esta afirmación basta observar que la relación

$$1 \cdot o + 0 \cdot a_1 + \ldots + 0 \cdot a_m = o$$

es una relación no trivial cualesquiera que sean $a_1, \ldots, a_m$.

La definición que hemos dado de dependencia lineal de un sistema presupone que dicho sistema contiene un número finito de vectores. Sin embargo, resulta necesario, con frecuencia, considerar también sistemas infinitos. Diremos que un sistema infinito de vectores es *linealmente dependiente*, si resulta linealmente dependiente alguna parte suya finita. El siguiente ejemplo muestra que pueden existir sistemas infinitos linealmente independientes de vectores. Consideremos el conjunto de todos los polinomios en letra λ con coeficientes de un cuerpo K. Respecto a las operaciones de adición y de multiplicación por números de K, estos polinomios forman, evidentemente, un espacio lineal. Los polinomios

$$1, \lambda, \lambda^2, \ldots, \lambda^m, \ldots$$

constituyen un sistema linealmente independiente en este espacio. Efectivamente, cualquier parte finita suya

$$\lambda^{m_1}, \lambda^{m_2}, \ldots, \lambda^{m_k}$$

$$(0 \leqslant m_1 < m_2 < \ldots < m_k)$$

es linealmente independiente, ya que de

$$\alpha_1 \lambda^{m_1} + \alpha_2 \lambda^{m_2} + \ldots + \alpha_k \lambda^{m_k} = o$$

se deduce que $\alpha_1 = \alpha_2 = \ldots = \alpha_k = 0$.

Consideremos un espacio lineal arbitrario $\mathfrak{L}$. Si un vector a de este espacio puede ser representado en la forma

$$a = \alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_m a_m$$

se dice que *a se expresa linealmente en términos de* $a_1, a_2, \ldots, a_m$ o que *a depende linealmente de* $a_1, \ldots, a_m$.

Si los vectores $a_1, \ldots, a_m$ *se expresan linealmente en términos de* $b_1, \ldots, b_n$ *y* $b_1, \ldots, b_n$ *se expresan linealmente en téminos de* $c_1, \ldots, c_p$, *los vectores* $a_1, \ldots a_m$ *se expresan linealmente en términos de* $c_1, \ldots, c_p$.

Para la demostración basta en las expresiones lineales de los vectores $a_1, \ldots, a_m$ en términos de $b_1, \ldots, b_n$ sustituir $b_1, \ldots b_n$ por sus expresiones en términos de $c_1, \ldots, c_p$ y reducir los términos semejantes.

TEOREMA 1. *Si un sistema de vectores no nulos* $a_1, a_2, \ldots, a_m$ *considerados en un orden determinado es linealmente dependiente, al menos uno de estos vectores puede ser expresado linealmente en términos de los anteriores. Recíprocamente, si uno de los vectores de esta sucesión se expresa linealmente en términos de los anteriores, el sistema es linealmente dependiente.*

Supongamos que entre los vectores $a_1, a_2, \ldots, a_m$ existe una relación no trivial

$$\alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_m a_m = o. \tag{1}$$

Sea α_k el último coeficiente diferente de cero. Si $k = 1$, la relación (1) se convierte en

$$\alpha_1 a_1 = o \quad (\alpha_1 \neq 0),$$

de donde $a_1 = o$ a pesar de haber aceptado que el sistema no contiene vectores nulos. Por consiguiente, $1 < k \leqslant m$ y la relación (1) puede ser representada en la forma

$$\alpha_1 a_1 + \ldots + \alpha_k a_k = o \quad (\alpha_k \neq 0),$$

de donde se tiene

$$a_k = -\alpha_k^{-1}\alpha_1 a_1 - \ldots - \alpha_k^{-1}\alpha_{k-1} a_{k-1}.$$

Con esto queda demostrada la primera parte del teorema. La afirmación recíproca es evidente.

Sea $\mathfrak{M}$ un conjunto de vectores de un espacio lineal $\mathfrak{L}$. Un sistema de vectores $a_1, a_2, \ldots$ de este conjunto se llama *sistema de generadores* de $\mathfrak{M}$, si todo vector de $\mathfrak{M}$ puede ser expresado linealmente en términos de un número finito de los vectores $a_1, a_2, \ldots$ Un sistema linealmente independiente de generadores de $\mathfrak{M}$ se llama *base* del conjunto $\mathfrak{M}$. Por ejemplo, en el espacio de todos los polinomios en λ los polinomios

$$1, \lambda, \lambda^2, \ldots, \lambda^m, \ldots \tag{2}$$

constituyen una base ya que estos polinomios, como hemos visto, son linealmente independientes y, por otro lado, todo polinomio es de la forma

$$\alpha_0 + \alpha_1\lambda + \alpha_2\lambda^2 + \ldots + \alpha_m\lambda^m,$$

es decir, se expresa linealmente en términos de $1, \lambda, \ldots, \lambda^m$.

LEMA *Si un sistema de generadores* $a_1, a_2, \ldots$ *de un conjunto* $\mathfrak{M}$ *contiene un elemento* a_i *que puede ser expresado linealmente en términos de los demás generadores, entonces suprimiendo* a_i *en el sistema de generadores se obtiene de nuevo un sistema de generadores de* $\mathfrak{M}$.

En efecto, por hipótesis todo vector de $\mathfrak{M}$ es una combinación lineal de un número finito de generadores. Sustituyendo en estas

combinaciones el vector a_i por su expresión en términos de los demás generadores obtendremos para cualquier elemento de $\mathfrak{M}$ una expresión en términos de los generadores diferentes de a_i que es lo que se quería demostrar.

TEOREMA 2. *De todo sistema de generadores de un espacio* $\mathfrak{L}$ *se puede extraer una base de este espacio.*

Supongamos que $\mathfrak{L}$ tiene un sistema finito de generadores $a_1, a_2, \ldots, a_s$. Omitiendo en este sistema todos los vectores que se expresan linealmente en términos de los anteriores, obtendremos un sistema de vectores $a_{i_1}, a_{i_2}, \ldots a_{i_n}$ que, de acuerdo con el lema, será aun un sistema de generadores del espacio $\mathfrak{L}$. Puesto que ningún vector de este último sistema puede ser expresado linealmente en términos de los anteriores, este sistema es, en virtud del teorema 1, linealmente independiente y, por lo tanto, es una base del espacio.

Hemos demostrado el teorema 2 aceptando que el sistema de generadores contiene un número finito de elementos. Sin embargo, este teorema es válido también en el caso de un sistema infinito de generadores. Para la demostración es suficiente disponer los generadores en una sucesión transfinita $a_1, a_2, \ldots, a_\omega, a_{\omega+1}, \ldots$ y omitir en ella todos los elementos que se expresan linealmente en términos de los anteriores.

Sea $a_1, a_2, \ldots$ una base del espacio $\mathfrak{L}$. Según la definición de una base, todo vector de $\mathfrak{L}$ se expresa linealmente en términos de un número finito de vectores básicos. Demostremos que esta representación es única. En efecto, sea

$$a = \alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_s a_s$$

y sea

$$a = \beta_1 a_1 + \beta_2 a_2 + \ldots + \beta_s a_s.$$

Restando obtenemos

$$o = (\alpha_1 - \beta_1) a_1 + (\alpha_2 - \beta_2) a_2 + \ldots + (\alpha_s - \beta_s) a_s.$$

Como $a_1, a_2, \ldots$ son linealmente independientes, se tiene $\alpha_1 - \beta_1 = 0$, $\alpha_2 - \beta_2 = 0, \ldots, \alpha_s - \beta_s = 0$, es decir, $\alpha_1 = \beta_1, \ldots, \alpha_s = \beta_s$, que es lo que se quería demostrar.

Si un espacio lineal $\mathfrak{L}$ tiene al menos una base compuesta por un número finito de elementos, se dice que $\mathfrak{L}$ es *de dimensión finita*. Con más detalle, $\mathfrak{L}$ se llama de dimensión finita si se puede escoger en $\mathfrak{L}$ un sistema linealmente independiente finito de vectores, tal, que todos los vectores de $\mathfrak{L}$ se expresan linealmente en términos de este sistema.

El concepto de base no se puede aplicar al espacio nulo. Sin embargo, aceptaremos que el espacio nulo es también de dimensión finita.

TEOREMA 3. *Todas las bases de un espacio lineal no nulo* $\mathfrak{L}$ *de dimensión finita constan de un mismo número finito de vectores. Este*

número se llama dimensión del espacio $\mathfrak{L}$. *La dimensión del espacio nulo es, por definición, el número cero.*

Según la definición, en $\mathfrak{L}$ existe una base que consta de un número finito de vectores. Sea

$$a_1, a_2, \ldots, a_n \tag{3}$$

esta base. Demostremos que el número de vectores de otra base cualquiera

$$x_1, x_2, \ldots, x_m, \ldots \tag{4}$$

no puede ser mayor que n. En efecto, consideremos el sistema

$$x_1, a_1, a_2, \ldots, a_n. \tag{5}$$

Puesto que el sistema (3) es un sistema de generadores de $\mathfrak{L}$, es decir, todo vector de $\mathfrak{L}$ se expresa linealmente en términos de los vectores (3), también el sistema (5) tendrá esta propiedad. Sin embargo, el sistema (5) es linealmente dependiente, ya que su primer vector x_1 puede ser expresado linealmente en términos de los restantes. Aplicando a la sucesión (5) el teorema 3, vemos que uno de los vectores de esta sucesión, digamos a_i, debe expresarse linealmente en términos de los anteriores. Omitiendo en (5) el vector a_i, obtendremos una sucesión nueva

$$x_1, a'_1, \ldots, a'_{n-1}. \tag{6}$$

donde $a'_1, \ldots, a'_{n-1}$ representan aquellos de los vectores $a_1, \ldots, a_n$ que hemos conservado. En virtud del lema, el sistema (6) será aun un sistema de generadores de $\mathfrak{L}$. Consideremos ahora la sucesión

$$x_2, x_1, a'_1, \ldots, a'_{n-1}. \tag{7}$$

Esta sucesión es linealmente dependiente debido a que el vector x_2 se expresa linealmente en términos de sus elementos restantes. Por consiguiente, de acuerdo con el teorema 1 uno de los vectores de este sistema debe expresarse linealmente en términos de *los anteriores*. Este vector debe ser sólo uno de los vectores $a'_1, \ldots, a'_{n-1}$, ya que x_2 y x_1 son, por hipótesis, linealmente independientes. Omitiéndolo de la sucesión (7), obtenemos la sucesión

$$x_2, x_1, a''_1, \ldots, a''_{n-2}, \tag{8}$$

donde $a''_1, \ldots, a''_{n-2}$ son aquellos de los vectores $a'_1, \ldots, a'_{n-1}$ que hemos conservado. Puesto que la sucesión (7) era un sistema de generadores de $\mathfrak{L}$, también la sucesión (8) será, en virtud del lema, un sistema de generadores de $\mathfrak{L}$. Agregando ahora a la sucesión (8) el vector x_3, omitiremos de la nueva sucesión el vector que se expresa linealmente en términos de los anteriores, etc. Si el número de vectores x_i fuese mayor que n, al cabo de n pasos obtendríamos

una sucesión

$$x_n, x_{n-1}, \ldots, x_2, x_1 \tag{9}$$

que no contendría los vectores $a_1, \ldots, a_n$ y que sería un sistema de generadores de $\mathfrak{L}$. Esto significaría que todos los vectores del espacio $\mathfrak{L}$ podrían ser expresados linealmente en términos de los vectores (9). En particular, también el vector x_{n+1} podría ser expresado linealmente en términos de los vectores (9), lo que estaría en contradicción con la independencia lineal de los vectores $x_1, x_2, \ldots$ Por consiguiente la base (4) no puede contener más vectores que la base (3), es decir, todas las bases de $\mathfrak{L}$ constan de un número *finito* de vectores. Por otro lado, la base (3) ha sido escogida arbitrariamente. Por lo tanto, con nuestro razonamiento queda demostrado que el número de vectores de una base no puede ser inferior al número de vectores de otra base, es decir, todas las bases de $\mathfrak{L}$ tienen un mismo número de vectores.

TEOREMA 4. *Cualquiera que sea un sistema linealmente independiente $a_1, \ldots, a_m$ de vectores de un espacio $\mathfrak{L}$ de dimensión finita, se puede encontrar en $\mathfrak{L}$ unos vectores $a_{m+1}, \ldots, a_n$ tales, que el sistema $a_1, \ldots, a_m, a_{m+1}, \ldots, a_n$ sea una base de $\mathfrak{L}$.*

DEMOSTRACIÓN. Escojamos en $\mathfrak{L}$ una base cualquiera $x_1, x_2, \ldots, x_n$ y consideremos la sucesión

$$a_1, a_2, \ldots, a_m, x_1, x_2, \ldots, x_n. \tag{10}$$

Omitamos ahora en esta sucesión todos aquellos vectores que se expresan linealmente en términos de los anteriores. Puesto que $a_1, a_2, \ldots a_m$ son linealmente independientes, no será omitido ninguno de ellos y el sistema que resulte será de la forma

$$a_1, \ldots, a_m, x_{i_1}, \ldots, x_{i_k}. \tag{11}$$

Debido al teorema 1 este sistema es linealmente independiente. Por otra parte, todos los vectores del espacio $\mathfrak{L}$ podían ser expresados linealmente en términos del sistema (10). En virtud del lema, la misma propiedad la debe tener también el sistema (11). Por consiguiente, el sistema (11) es una base del espacio $\mathfrak{L}$ y $x_{i_1}, x_{i_2}, \ldots, x_{i_k}$ son los vectores que queríamos encontrar.

El teorema 4 puede ser enunciado también en esta forma: *todo sistema linealmente independiente de vectores del espacio $\mathfrak{L}$ o bien es una base o bien forma parte de una base de $\mathfrak{L}$.*

Supongamos que $\mathfrak{L}$ es de dimensión n. Entonces, toda base del espacio $\mathfrak{L}$ contiene n vectores y, por consiguiente, el número de vectores en cualquier sistema linealmente independiente de $\mathfrak{L}$ es o menor de n o igual a n. En el último caso el sistema debe ser una base de $\mathfrak{L}$. En particular, el número máximo de vectores linealmente independientes de $\mathfrak{L}$ es igual a n, es decir, es igual a la dimensión de $\mathfrak{L}$.

Hemos obtenido el teorema siguiente.

TEOREMA 5 *Todo sistema de $n+1$ vectores de un espacio lineal $\mathfrak{L}$ de n dimensiones es linealmente dependiente. Cualesquiera n vectores linealmente independientes de este espacio constituyen una base del mismo. El número máximo de vectores linealmente independientes del espacio $\mathfrak{L}$ es igual a la dimensión de este espacio.*

Para concluir, observemos que los teoremas 3 y 4, que hemos enunciado para el caso de espacios de dimensión finita, tienen lugar también en los espacios de dimensión infinita. Solamente en lugar del número de vectores de una base habrá que considerar la potencia del conjunto de los vectores que la componen y, respectivamente, la dimensión de un espacio deberá entenderse como la potencia del conjunto de vectores que forman una base cualquiera de este espacio. En cuanto al teorema 5, solamente la última de sus afirmaciones puede ser extendida directamente al caso de dimensión infinita. En este libro estudiaremos las propiedades de los espacios de dimensión finita. Por esto, *en lo sucesivo un espacio lineal se comprenderá, siempre que no se diga lo contrario, como un espacio lineal de dimensión finita.*

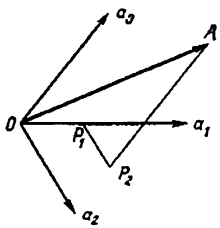


Fig. 1.

Determinemos la dimensión de los espacios considerados en el p.4.1. Sea $\mathfrak{R}$ el espacio habitual de segmentos orientados que parten de un punto O. Como de costumbre, diremos que estos segmentos son vectores. Demostremos que *cualesquiera tres vectores a_1, a_2 y a_3 del espacio $\mathfrak{R}$, que parten del punto O y que no pertenecen a un mismo plano, constituyen una base de $\mathfrak{R}$.* En efecto, los vectores a_1, a_2 y a_3 son linealmente independientes, ya que de lo contrario uno de ellos, digamos a_3, debería expresarse linealmente en términos de los otros dos. Sin embargo, la relación $a_3 = \alpha_1 a_1 + \alpha_2 a_2$ significa que a_3 es la diagonal del paralelogramo construido a partir de los vectores $\alpha_1 a_1$ y $\alpha_2 a_2$. Puesto que $\alpha_1 a_1$ y $\alpha_2 a_2$ se hallan en el plano $a_1 O a_2$, también a_3 tendría que pertenecer al mismo plano, lo que estaría en contradicción con la hipótesis. Por otra parte, todo vector $\overrightarrow{OA}$ del espacio $\mathfrak{R}$ puede ser expresado linealmente en términos de a_1, a_2 y a_3 (fig. 1):

$$\overrightarrow{OA} = \overrightarrow{OP}_1 + \overrightarrow{P_1P_2} + \overrightarrow{P_2A} = \alpha_1 a_1 + \alpha_2 a_2 + \alpha_3 a_3,$$

donde α_i es la relación entre la longitud del segmento $P_{i-1}P_i$ ($P_0 = O$, $P_3 = A$) y la longitud de a_i tomado con el signo adecuado. Por consiguiente, a_1, a_2 y a_3 es una base del espacio $\mathfrak{R}$ y $\mathfrak{R}$ es de dimensión tres.

Análogamente se demuestra que el conjunto de vectores, que parten de un punto O y que pertenecen a un plano que pasa por O, es un espacio lineal de dos dimensiones y que el conjunto de vectores, que parten de O y que pertenecen a una recta que pasa por O, es un espacio de una dimensión.

Consideremos el espacio de filas de longitud n formadas por elementos de un cuerpo K. Sea

$$e_1 = [1, 0, \ldots, 0],$$
$$e_2 = [0, 1, \ldots, 0],$$
$$\ldots\ldots\ldots\ldots$$
$$e_n = [0, 0, \ldots, 1].$$

Multiplicando sucesivamente estas filas por números arbitrarios $\alpha_1, \alpha_2, \ldots, \alpha_n$ y sumando, obtenemos

$$\alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_n e_n = [\alpha_1, \alpha_2, \ldots, \alpha_n].$$

Es decir, una fila arbitraria $[\alpha_1, \ldots, \alpha_n]$ se expresa linealmente en términos de $e_1, \ldots, e_n$. Sin embargo, el sistema $e_1, \ldots, e_n$ es linealmente independiente, debido a que la relación

$$\alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_n e_n = o$$

implica

$$[\alpha_1, \alpha_2, \ldots, \alpha_n] = [0, 0, \ldots, 0],$$

de donde se tiene $\alpha_1 = \alpha_2 = \ldots = \alpha_n = 0$. Por ello, $e_1, e_2, \ldots, e_n$ es una base del espacio considerado y, por consiguiente, su dimensión es n.

Finalmente, sea $\mathfrak{L}$ el espacio de funciones definidas sobre un conjunto finito $\mathfrak{M}$ y con valores en un cuerpo K. Sean $m_1, m_2, \ldots, m_s$ los elementos de $\mathfrak{M}$. Según el p. 4.1, toda función f de $\mathfrak{L}$ puede ser representada en la forma

$$f = \alpha_1 [m_1] + \alpha_2 [m_2] + \ldots + \alpha_s [m_s],$$

donde $[m_i]$ es la función igual a 1 sobre m_i e igual a 0 sobre los demás elementos del conjunto $\mathfrak{M}$. Es decir, f se expresa linealmente en términos de $[m_1], \ldots, [m_s]$. Por otro lado, la igualdad

$$\alpha_1 [m_1] + \alpha_2 [m_2] + \ldots + \alpha_s [m_s] = o$$

significa que todos los valores de la función que figura en el primer miembro son iguales a cero y, por consiguiente, $\alpha_1 = \ldots = \alpha_s = 0$. Vemos de esta forma que $[m_1], \ldots, [m_s]$ es una base del espacio $\mathfrak{L}$ y que su dimensión es igual a s.

Hemos considerado espacios lineales es decir, módulos (unitarios) sobre un cuerpo K. Ejemplos muy sencillos permiten ver que para los módulos sobre *anillos arbitrarios*, o incluso sobre anillos conmutativos arbitrarios, los teoremas 1 y 2 dejan de tener lugar. Sea, en particular, K el anillo de los números enteros corrientes y sea $\mathfrak{L}$

el conjunto de estos mismos números enteros en el que la operación de adición está definida del modo corriente y la operación de multiplicación de «números» de K por «vectores» de $\mathfrak{L}$ está definida como la multiplicación habitual de números enteros. Convendremos en indicar los números enteros que representan los vectores de $\mathfrak{L}$ con cifras gruesas, mientras que los números de K se indicarán con cifras corrientes. Puesto que para cualquier vector m es válida la fórmula

$$m = m \cdot \mathbf{1},$$

nuestro módulo numérico $\mathfrak{L}$ resulta ser generado por el vector 1. Los vectores 2 y 3 son linealmente dependientes, ya que

$$2 \cdot \mathbf{3} + (-3) \cdot \mathbf{2} = \mathbf{0}.$$

Sin embargo, para cualquier $\alpha \in K$ se tiene $\alpha \cdot \mathbf{3} \neq \mathbf{2}$ y $\alpha \cdot \mathbf{2} \neq \mathbf{3}$, es decir, en $\mathfrak{L}$ ninguno de los vectores 2 y 3 se expresa linealmente en términos del otro. Las propiedades de los módulos sobre anillos arbitrarios son estudiadas sistemáticamente en la teoría de los módulos, estrechamente ligada a la teoría de los números. Para nosotros, en cambio, tendrán interés solamente las propiedades de los espacios vectoriales y tocaremos la teoría de los módulos sobre anillos sólo cuando esto sea necesario para la teoría de los espacios vectoriales.

4.3. Isomorfismo. El concepto de espacio lineal tiene dos facetas esencialmente diferentes. En primer lugar, un espacio lineal es un conjunto de ciertas entes que se denominan vectores y, en segundo lugar, en un espacio lineal actúan las operaciones de adición y de multiplicación por número. Por esto, o bien podemos limitarnos a estudiar qué es lo que representan los vectores y cuáles son la naturaleza y las propiedades de los mismos, o bien podemos tomar otro punto de vista y estudiar las propiedades de las operaciones indicadas independientemente de la naturaleza de los elementos con los cuales se efectúan estas operaciones. En lo sucesivo nos interesarán solamente las propiedades del segundo género. Por ello, dos espacios de la misma estructura respecto a las operaciones de adición y de multiplicación por número se considerará que tienen las mismas propiedades o que son *isomorfos*. Con más precisión el concepto de *isomorfismo* puede enunciarse del modo siguiente:

Dos espacios lineales sobre un mismo cuerpo de coeficientes se llaman isomorfos, si se puede establecer una correspondencia biyectiva entre sus elementos, tal que a la suma de vectores del primer espacio corresponda la suma de los vectores correspondientes del segundo espacio y al producto de un número por un vector del primer espacio corresponda el producto de este mismo número por el vector correspondiente del segundo espacio.

Toda correspondencia biyectiva que posee las propiedades indicadas se llama *isomorfa* o *isomorfismo*. Consideremos las propiedades elementales de los isomorfismos.

En una correspondencia isomorfa el vector nulo corresponde al vector nulo. En efecto, supongamos que en una aplicación isomorfa de un espacio lineal $\mathfrak{L}$ sobre otro espacio lineal $\mathfrak{L}_1$ el vector a de $\mathfrak{L}$ corresponde al vector a_1 de $\mathfrak{L}_1$. Entonces, según la definición de un isomorfismo, el producto $0 \cdot a$ debe corresponder al producto $0 \cdot a_1$, es decir, el vector nulo del primer espacio debe transformarse en el vector nulo del segundo espacio.

En una aplicación isomorfa un sistema de generadores del primer espacio se transforma en un sistema de generadores del segundo espacio.

Efectivamente, sean $a_1, a_2, \ldots, a_s$ unos generadores del primer espacio y sean $b_1, b_2, \ldots, b_s$ los vectores que les corresponden en el segundo espacio. Tomemos en el segundo espacio un vector arbitrario b y consideremos el vector a del primer espacio que le corresponde. Por hipótesis, el vector a puede ser representado en la forma

$$a = \alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_s a_s.$$

Según la definición de una aplicación isomorfa, la suma $\alpha_1 a_1 + \ldots + \alpha_s a_s$ debe transformarse en la suma $\alpha_1 b_1 + \ldots + \alpha_s b_s$ y, por consiguiente, el vector b debe coincidir con la suma $\alpha_1 b_1 + \ldots + \alpha_s b_s$, es decir, los vectores $b_1, \ldots, b_s$ constituyen un sistema de generadores del segundo espacio.

En un isomorfismo los vectores linealmente independientes se transforman en vectores linealmente independientes.

En efecto, supongamos que los vectores linealmente independientes $a_1, a_2, \ldots, a_m$ del primer espacio se transforman en los vectores $b_1, b_2, \ldots, b_m$ del segundo espacio. Supongamos que entre los últimos existe una relación de tipo

$$\beta_1 b_1 + \beta_2 b_2 + \ldots + \beta_m b_m = o_1.$$

Según la definición de un isomorfismo, al primer miembro de esta igualdad corresponde en el primer espacio el vector $\beta_1 a_1 + \beta_2 a_2 + \ldots + \beta_m a_m$ y al vector nulo o_1 corresponde en el primer espacio el vector nulo o. Por consiguiente,

$$\beta_1 a_1 + \beta_2 a_2 + \ldots + \beta_m a_m = o.$$

Puesto que los vectores $a_1, \ldots, a_m$ son linealmente independientes se tiene

$$\beta_1 = \beta_2 = \ldots = \beta_m = 0,$$

es decir, los vectores $b_1, \ldots, b_m$ son linealmente independientes.

De las dos propiedades últimas se deduce directamente que *en un isomorfismo una base de un espacio lineal se transforma de nuevo en una base de un espacio lineal* y, por consiguiente, *los espacios lineales isomorfos tienen la misma dimensión.*

La afirmación recíproca es también válida: *si dos espacios lineales sobre un mismo cuerpo de coeficientes tienen la misma dimensión, son isomorfos.*

Para la demostración tomemos una base cualquiera en cada uno de los espacios dados, por ejemplo, $a_1, a_2, \ldots, a_n$ y $b_1, b_2, \ldots, b_n$. Diremos que los vectores

$$a = \alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_n a_n$$

y

$$b = \beta_1 b_1 + \beta_2 b_2 + \ldots + \beta_n b_n$$

son *correspondientes*, si $\alpha_1 = \beta_1, \ldots, \alpha_n = \beta_n$. Puesto que todo vector de un espacio se expresa linealmente en términos de una base de un modo único, nuestra correspondencia es biyectiva. Sean ahora

$$a = \alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_n a_n$$

y

$$b = \alpha_1 b_1 + \alpha_2 b_2 + \ldots + \alpha_n b_n$$

dos vectores correspondientes. Entonces se tiene

$$\alpha a = \alpha\alpha_1 a_1 + \alpha\alpha_2 a_2 + \ldots + \alpha\alpha_n a_n$$

y

$$\alpha b = \alpha\alpha_1 b_1 + \alpha\alpha_2 b_2 + \ldots + \alpha\alpha_n b_n.$$

Puesto que en estas descomposiciones coinciden los coeficientes respectivos, los vectores αa y αb serán correspondientes, es decir, en nuestra correspondencia el producto de un número por un vector se transforma en el producto del mismo número por el vector correspondiente. Análogamente se demuestra que la suma de vectores se transforma en la suma de vectores correspondientes. Por esto, la correspondencia construida es un isomorfismo, que es lo que se quería demostrar.

Las propiedades de las correspondencias isomorfas que hemos enunciado muestran que fijado el cuerpo principal K todo espacio lineal queda determinado, salvo un isomorfismo, por su dimensión. Por esta razón, los espacios de filas de longitud n con elementos de un cuerpo K, donde $n = 1, 2, \ldots$, agotan, salvo isomorfismos, todos los espacios de dimensión finita sobre K. En particular, el espacio habitual de segmentos orientados es isomorfo al espacio de filas de longitud tres sobre el cuerpo de los números reales, el espacio de funciones definidas sobre un conjunto $\mathfrak{M}$, compuesto por s elementos, y con valores en un cuerpo K es isomorfo al espacio de filas de longitud s con elementos de K, etc.

Para concluir hagamos una observación más. En el Algebra general desempeñan un papel principal el concepto de álgebra de signatura dada y el concepto de isomorfismo de álgebras de signatura determinada. En la definición que hemos

dado del concepto de módulo no se ha indicado, desde el punto de vista formal, la signatura del módulo, es decir, no se han señalado las operaciones que se consideran como principales y respecto a las cuales se define el concepto de isomorfismo. Definiendo (al igual que para los espacios lineales) el concepto de isomorfismo de módulos, fijamos con ello el conjunto de las operaciones principales, aunque de un modo implícito. Indicando estas operaciones explícitamente, obtenemos la siguiente definición de módulo unitario que sólo en la forma difiere de la definición del p. 4.1.

Se llama *módulo unitario* sobre un anillo K con el elemento unidad 1 un álgebra cuya signatura se compone del símbolo $+$ de una operación binaria y de los símbolos $-$ y F_α $(\alpha \in K)$ de operaciones de una posición siempre que en este álgebra se cumplan las identidades

$$M_1\colon\ x+y=y+x;$$
$$M_2\colon\ x+(y+z)=(x+y)+z;$$
$$M_3\colon\ x+(y+(-y))=x;$$
$$M_4\colon\ F_\alpha(F_\beta(a))=F_{\alpha\beta}(a);$$
$$M_5\colon\ F_\alpha(a+b)=F_\alpha(a)+F_\alpha(b);$$
$$M_6\colon\ F_{\alpha+\beta}(a)=F_\alpha(a)+F_\beta(a);$$
$$M_7\colon\ F_1(a)=a.$$

Por consiguiente, la multiplicación de los elementos del conjunto principal (de los vectores) por cualquier elemento fijo $\alpha \in K$ se considera aquí como una operación principal independiente. Si el anillo principal K es infinito, la signatura del módulo es también infinita. Cambiando el anillo K, cambiamos también la signatura de la clase de módulos.

Existe, en general, otra forma de incluir la teoría de módulos en la teoría general de álgebras. Para ello los módulos se consideran como álgebras compuestas de dos conjuntos principales: el conjunto de los números y el conjunto de los vectores. La signatura consta ahora de las operaciones de adición de números, de paso al número opuesto, de adición de vectores, de paso al vector opuesto, de multiplicación de un número por un vector y de la operación 0-aria que despeja la unidad 1 (en total seis operaciones). Las identidades principales son en este caso las identidades que definen un anillo, las identidades que definen un grupo conmutativo y las identidades de 4° a 7° del p. 4.1.

Este nuevo concepto de módulo es diferente del anterior. Todos los módulos tienen, en el sentido nuevo, una misma signatura (las seis operaciones indicadas) y por esto resulta posible preguntar si son o no isomorfos unos módulos definidos sobre distintos anillos. Estos isomorfismos nuevos suelen llamarse, en diferencia de los definidos anteriormente, *antiisomorfismos*. En este orden de ideas se definen también los *antiautomorfismos* y los *antiendomorfismos* y otros conceptos análogos.

Ejemplos y problemas

1. Demuéstrese que la dimensión del espacio de todos los polinomios en una variable de grado no mayor que n es igual a $n+1$.
2. Los polinomios homogéneos en dos variables de grado n constituyen un espacio lineal de dimensión $n+1$.
3. ¿Cuál es la dimensión del espacio de los polinomios homogéneos en k variables de grado n?
4. Las matrices de m filas y de n columnas formadas por elementos de un cuerpo dado K constituyen un espacio lineal respecto a las operaciones usuales matriciales de adición y de multiplicación por número. Demuéstrese que las matrices en las que un elemento es igual a la unidad y todos los demás ele-

mentos son iguales a cero, constituyen una base de este espacio y que, por consiguiente, la dimensión de este espacio es igual a mn.

5. Las matrices simétricas, así como las matrices antisimétricas, de orden n formadas por elementos de un cuerpo K constituyen unos espacios lineales sobre K. Demuéstrese que las dimensiones de estos espacios son iguales a $\frac{n(n+1)}{2}$ y a $\frac{n(n-1)}{2}$, respectivamente.

§ 5. Coordenadas

5.1. Coordenadas de un vector. En el parágrafo anterior hemos considerado las propiedades generales elementales de los espacios lineales. Sin embargo, en las aplicaciones, además de conocer las propiedades generales, es importante saber definir los vectores en términos de números y poder reducir las operaciones vectoriales a operaciones con números. Este problema se resuelve introduciendo las coordenadas de un espacio vectorial.

Toda base de un espacio lineal $\mathfrak{L}$, cuyos vectores se toman en un orden determinado, se llamará *base de coordenados* o *sistema de coordenadas* de $\mathfrak{L}$. Por consiguiente, si

$$a_1, a_2, \ldots, a_n \tag{1}$$

es un sistema de coordenadas de $\mathfrak{L}$, estos mismos vectores, pero tomados en otro orden, representarán otro sistema de coordenadas de $\mathfrak{L}$. Hemos visto que todo vector a de $\mathfrak{L}$ puede ser representado unívocamente en la forma siguiente:

$$a = \alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_n a_n. \tag{2}$$

Los números $\alpha_1, \ldots, \alpha_n$ se llaman *coordenadas del vector* a en el sistema de coordenadas (1). La fila $[\alpha_1, \alpha_2, \ldots, \alpha_n]$ compuesta por las coordenadas del vector a, tomadas en un orden adecuado, se llama *fila de coordenadas* y se indica por $[a]$. Por consiguiente, una vez escogido en el espacio un sistema de coordenadas determinado, a todo vector corresponde una fila de coordenadas y, viceversa, para toda fila de longitud n se obtiene con la fórmula (2) un vector determinado a, para el cual esta fila es su fila de coordenadas.

Supongamos que $[\alpha_1, \ldots, \alpha_n]$ y $[\beta_1, \ldots, \beta_n]$ son las filas de coordenadas de los vectores a y b, es decir,

$$a = \alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_n a_n$$

y

$$b = \beta_1 a_1 + \beta_2 a_2 + \ldots + \beta_n a_n.$$

Es evidente que

$$\alpha a = (\alpha\alpha_1) a_1 + (\alpha\alpha_2) a_2 + \ldots + (\alpha\alpha_n) a_n$$

y

$$a+b=(\alpha_1+\beta_1)a_1+(\alpha_2+\beta_2)a_2+\ldots+(\alpha_n+\beta_n)a_n.$$

Empleando las reglas de operaciones con filas, estas igualdades pueden ser representadas en la forma

$$[\alpha a]=\alpha[a] \quad \text{y} \quad [a+b]=[a]+[b].$$

Por consiguiente, *la fila de coordenadas de una suma de vectores es igual a la suma de las filas de coordenadas de los sumandos y la fila de coordenadas del producto de un número por un vector es igual al producto de este número por la fila de coordenadas del vector.*

Este resultado puede ser interpretado de la forma siguiente. Sea $\mathfrak{L}$ un espacio lineal de dimensión n sobre un cuerpo K. Sea $\mathfrak{L}_n$ el espacio de filas de longitud n formadas por elementos de K. Tomemos en $\mathfrak{L}$ un sistema de coordenadas determinado y pongamos en correspondencia a todo vector de $\mathfrak{L}$ su fila de coordenadas. Nuestro resultado significa que esta correspondencia es un isomorfismo entre $\mathfrak{L}$ y $\mathfrak{L}_n$. En particular, de aquí se desprende que los vectores linealmente independientes tienen filas de coordenadas linealmente independientes y que toda relación de dependencia lineal entre los vectores dados tiene lugar también para las filas de coordenadas de los mismos.

En un mismo espacio $\mathfrak{L}$ existen diferentes sistemas de coordenadas. Por esto surge la pregunta: ¿cómo varían las coordenadas de un vector al cambiar un sistema de coordenadas por otro? Para resolver este problema, tomemos en $\mathfrak{L}$ dos sistemas de coordenadas $a_1, a_2, \ldots, a_n$ y $a'_1, a'_2, \ldots, a'_n$ cualesquiera. Puesto que los vectores $a_1, \ldots, a_n$ constituyen una base de $\mathfrak{L}$, los vectores $a'_1, \ldots, a'_n$ deben expresarse linealmente en términos de $a_1, \ldots, a_n$. Sean

$$\begin{aligned}
a'_1 &= \tau_{11}a_1+\tau_{12}a_2+\ldots+\tau_{1n}a_n,\\
a'_2 &= \tau_{21}a_1+\tau_{22}a_2+\ldots+\tau_{2n}a_n,\\
&\cdots\cdots\cdots\\
a'_n &= \tau_{n1}a_1+\tau_{n2}a_2+\ldots+\tau_{nn}a_n.
\end{aligned}$$

estas expresiones. La matriz

$$T=\begin{bmatrix}\tau_{11} & \tau_{12} & \ldots & \tau_{1n}\\ \tau_{21} & \tau_{22} & \ldots & \tau_{2n}\\ \cdots & \cdots & \cdots & \cdots\\ \tau_{n1} & \tau_{n2} & \ldots & \tau_{nn}\end{bmatrix}$$

se llama *matriz del cambio* del sistema de coordenadas $a_1, \ldots, a_n$ por el sistema de coordenadas $a'_1, \ldots, a'_n$. Tomemos un vector a cualquiera y sean $[\alpha_1, \ldots, \alpha_n]$ y $[\alpha'_1, \ldots, \alpha'_n]$ sus filas de coordenadas en los sistemas antiguo y nuevo de coordenadas. Es decir,

tenemos $$a = \alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_n a_n$$

y

$$a = \alpha'_1 a'_1 + \alpha'_2 a'_2 + \ldots + \alpha'_n a'_n.$$

Tomando en la segunda de estas igualdades en lugar de los vectores $a'_1, \ldots, a'_n$ sus expresiones en términos de $a_1, \ldots, a_n$, obtenemos

$$a = (\alpha'_1 \tau_{11} + \alpha'_2 \tau_{21} + \ldots + \alpha'_n \tau_{n1}) a_1 + (\alpha'_1 \tau_{12} + \ldots + \alpha'_n \tau_{n2}) a_2 + \ldots,$$

es decir

$$\begin{aligned} \alpha_1 &= \alpha'_1 \tau_{11} + \alpha'_2 \tau_{21} + \ldots + \alpha'_n \tau_{n1}, \\ \alpha_2 &= \alpha'_1 \tau_{12} + \alpha'_2 \tau_{22} + \ldots + \alpha'_n \tau_{n2}, \\ &\ldots\ldots\ldots\ldots\ldots \\ \alpha_n &= \alpha'_1 \tau_{1n} + \alpha'_2 \tau_{2n} + \ldots + \alpha'_n \tau_{nn}. \end{aligned}$$

Estas igualdades son precisamente las fórmulas de *transformación de coordenadas* que buscábamos. Observando que la expresión para α_i:

$$\alpha_i = \alpha'_1 \tau_{1i} + \alpha'_2 \tau_{2i} + \ldots + \alpha'_n \tau_{ni}$$

representa el producto de la fila de coordenadas $[\alpha'_1, \ldots, \alpha'_n]$ por la i-ésima columna de la matriz T, vemos que todo el sistema de las fórmulas de transformación de coordenadas puede ser representado brevemente en la forma matricial:

$$[\alpha_1, \ldots, \alpha_n] = [\alpha'_1, \ldots, \alpha'_n]\, T.$$

Hemos obtenido la siguiente regla:

REGLA DE TRANSFORMACIÓN DE COORDENADAS. *La fila de coordenadas antigua de un vector es igual a la nueva multiplicada por la matriz del cambio.*

Demostremos el siguiente lema que en muchas ocasiones resulta útil.

LEMA. *Sean A y B dos matrices cuadradas de orden n formadas por elementos de un cuerpo K. Si para cualquier fila $[\xi_1, \ldots, \xi_n]$ de elementos de K resulta*

$$[\xi_1, \ldots, \xi_n]\, A = [\xi_1, \ldots, \xi_n]\, B, \tag{3}$$

se tiene $A = B$.

En efecto, sean α_{ij} los elementos de la matriz A y sean β_{ij} los elementos de la matriz B ($i, j = 1, \ldots, n$); entonces cualquiera que sea i para $\xi_j = 1$ y $\xi_j = 0$ ($i \neq j$) de la igualdad (3) resulta que $\alpha_{ik} = \beta_{ik}$ ($i, k = 1, \ldots, n$), que es lo que se quería demostrar.

Consideremos ahora en un espacio lineal Ω de n dimensiones dos sistemas de coordenadas $a_1, \ldots, a_n$ y $a'_1, \ldots, a'_n$. Podemos expresar o bien $a'_1, \ldots, a'_n$ en términos de $a_1, \ldots, a_n$:

$$\begin{aligned} a'_1 &= \tau_{11} a_1 + \ldots + \tau_{1n} a_n, \\ &\ldots\ldots\ldots\ldots\ldots \\ a'_n &= \tau_{n1} a_1 + \ldots + \tau_{nn} a_n, \end{aligned} \tag{4}$$

o bien, al contrario, $a_1, \ldots, a_n$ en términos de $a'_1, \ldots, a'_n$:

$$a_1 = \sigma_{11}a'_1 + \ldots + \sigma_{1n}a'_n,$$
$$\ldots\ldots\ldots\ldots\ldots$$
$$a_n = \sigma_{n1}a'_1 + \ldots + \sigma_{nn}a'_n.$$

La matriz T, formada por los coeficientes τ_{ij}, es la matriz del cambio del sistema $a_1, \ldots, a_n$ por el sistema $a'_1, \ldots, a'_n$ y la matriz S, formada por los coeficientes σ_{ij}, es la matriz del *cambio recíproco* del sistema $a'_1, \ldots, a'_n$ por el sistema $a_1, \ldots, a_n$. Sea $[x]$ la fila de coordenadas de un vector cualquiera x en el sistema de coordenadas $a_1, \ldots, a_n$ y sea $[x]_1$ la fila de coordenadas del vector x en el sistema $a'_1, \ldots, a'_n$. Considerando que $a_1, \ldots, a_n$ es el sistema antiguo y aplicando la regla de transformación de coordenadas, obtenemos

$$[x] = [x]_1 T. \tag{5}$$

Al contrario, considerando que el sistema antiguo es $a'_1, \ldots, a'_n$, tendremos

$$[x]_1 = [x]\, S.$$

Tomando aquí en lugar de $[x]$ su expresión (5), obtenemos

$$[x]_1 = [x]_1\, TS$$

y de forma análoga

$$[x] = [x]\, ST.$$

Aquí el vector x y, por consiguiente, sus filas de coordenadas $[x]_1$ y $[x]$ son arbitrarias. En virtud del lema, obtenemos entonces

$$E = TS = ST \quad \text{y} \quad S = T^{-1}.$$

Es decir, *la matriz de la transformación de coordenadas siempre posee la inversa* que es la matriz de la transformación recíproca de coordenadas.

Nuestros razonamientos permiten aclarar un detalle relacionado con las matrices inversas que hemos dejado sin consideración en el p. 2.3. Según el p. 2.3, una matriz cuadrada S se llama inversa de la matriz T, si se cumplen dos igualdades:

$$ST = E \quad \text{y} \quad TS = E.$$

Si S satisface sólo la primera o sólo la segunda de estas igualdades, se dice que es *inversa a la izquierda* y, respectivamente, *inversa a la derecha*.

TEOREMA *Si una matriz cuadrada T formada por elementos de un cuerpo K posee la matriz inversa a la izquierda o a la derecha S, la matriz S es simplemente la inversa de T.*

Tomemos un espacio $\mathfrak{L}$ de n dimensiones con una base cualquiera $a_1, \ldots, a_n$. Calculemos según las fórmulas (4) los vectores $a'_1, \ldots, a'_n$.

Demostremos que son linealmente independientes. Sea

$$\lambda_1 a'_1 + \ldots + \lambda_n a'_n = o.$$

Es obvio que esta relación equivale a la condición matricial

$$[\lambda_1, \ldots, \lambda_n] \cdot T = o. \tag{6}$$

Si T posee la matriz inversa a la derecha S, multiplicando a la derecha ambos miembros de la igualdad (6) por S y observando que $TS = E$, obtenemos $[\lambda_1, \ldots, \lambda_n] = o$. Por consiguiente, los vectores $a'_1, \ldots, a'_n$ son linealmente independientes y el sistema $a'_1, \ldots, a'_n$ puede ser considerado como un nuevo sistema de coordenadas de $\mathfrak{L}$, con la particularidad de que la matriz T será la matriz del cambio. La matriz del cambio posee la matriz inversa T^{-1}. Multiplicando la relación $TS = E$ a la izquierda por T^{-1}, obtenemos $S = T^{-1}$. Hemos demostrado que una matriz inversa a la derecha es simplemente la matriz inversa. Análogamente se demuestra que una matriz inversa a la izquierda es también simplemente la inversa.

De los razonamientos realizados se desprende, en particular, que toda matriz invertible de orden n es una matriz del cambio de determinados sistemas de coordenadas.

5.2. Rangos de matrices. Consideremos un espacio lineal $\mathfrak{L}$ de dimensión finita n sobre un cuerpo K. Tomemos en $\mathfrak{L}$ una base cualquiera $a_1, \ldots, a_n$ a título del sistema de coordenadas y supongamos que los vectores $x_1, \ldots, x_m$ tienen respectivamente las siguientes filas de coordenadas

$$[x_1] = [\alpha_{11}, \ldots, \alpha_{1n}],$$
$$\ldots\ldots\ldots\ldots$$
$$[x_m] = [\alpha_{m1}, \ldots, \alpha_{mn}].$$

Formemos a partir de estas filas la matriz

$$A = \begin{bmatrix} \alpha_{11} & \ldots & \alpha_{1n} \\ \ldots & \ldots & \ldots \\ \alpha_{m1} & \ldots & \alpha_{mn} \end{bmatrix}.$$

Sabemos ya que el número máximo de vectores linealmente independientes en el sistema $x_1, \ldots, x_m$ es igual al número máximo de filas linealmente independientes en el sistema $[x_1], \ldots, [x_m]$, es decir, coincide con el número máximo de filas linealmente independientes de la matriz A.

El número máximo de filas linealmente independientes de una matriz arbitraria A, formada por elementos de un cuerpo dado K, se llama *rango de la matriz* A, o, con más precisión, *rango según las filas*.

De este modo, el número máximo de vectores linealmente independientes entre los vectores $x_1, \ldots, x_m$ es igual al rango de la matriz formada por las filas de coordenadas de estos vectores.

Volvamos a considerar una matriz arbitraria A de m filas y de n columnas. Escojamos en A cualesquiera k filas y k columnas. Los elementos de la matriz A, que se hallan en los cruces de estas filas y columnas, tomadas en su orden natural, forman una matriz cuadrada de orden k, que se llama menor de orden k de la matriz A. Considerando todos los menores de primero, segundo, etc. órdenes, de la matriz A, podremos ver que una parte de los mismos serán matrices invertibles y otra parte, matrices no invertibles. El orden máximo de los menores invertibles de la matriz A se llama *rango* de la misma *según los menores* o *rango de menores*.

Los menores de primer orden son los elementos de la matriz A. Puesto que la matriz A se toma sobre un cuerpo, el hecho de que un menor de primer orden no sea invertible significa simplemente que este menor está formado por el número cero. Luego, una matriz A no posee menores invertibles si, y sólo si, está compuesta de ceros. En este caso se dice que el rango de menores de la matriz A es igual a cero. Observemos que el rango según las filas de la matriz nula es también igual a cero, ya que un sistema de vectores nulos no posee subsistemas linealmente independientes. Pretendemos demostrar ahora que el rango según las filas de una matriz arbitraria coincide con su rango de menores. Antes de obtener este resultado, consideremos las matrices cuadradas.

TEOREMA 1 *Para que una matriz cuadrada* $A = \|\alpha_{ij}\|_{nn}$, *formada por elementos de un cuerpo* K, *sea invertible es necesario y suficiente que sus filas sean linealmente independientes.*

Sea a_i $(i = 1, \ldots, n)$ la i-ésima fila de la matriz. Entonces toda relación de dependencia lineal

$$\lambda_1 a_1 + \ldots + \lambda_n a_n = o \tag{1}$$

entre las filas de la matriz A puede ser representada en la forma

$$[\lambda_1, \ldots, \lambda_n] \cdot A = o. \tag{2}$$

Siendo la matriz A invertible y multiplicando ambos miembros de la igualdad (2) por A^{-1}, obtenemos $[\lambda_1, \ldots, \lambda_n] = o$ y con esto queda demostrado que las condiciones del teorema 1 son necesarias.

Recíprocamente, supongamos que las filas $a_1, \ldots, a_n$ de la matriz A son linealmente independientes. Sea $\mathfrak{L}$ el espacio vectorial de todas las filas de longitud n formadas por elementos del cuerpo K. Las filas

$$\begin{aligned} e_1 &= [1, 0, \ldots, 0], \\ e_2 &= [0, 1, \ldots, 0], \\ &\ldots\ldots\ldots\ldots \\ e_n &= [0, 0, \ldots, 1] \end{aligned}$$

constituyen una base de $\mathfrak{L}$. Por otra parte, el espacio $\mathfrak{L}$ es de n dimensiones y, por hipótesis, los vectores $a_1, \ldots, a_n$ son linealmente independientes. Luego, los vectores $a_1, \ldots, a_n$ forman también

una base del espacio $\mathfrak{L}$. Considerando la base $e_1, \ldots, e_n$ como el sistema de coordenadas inicial en $\mathfrak{L}$ y la base $a_1, \ldots, a_n$ como un nuevo sistema de coordenadas, vemos que A es la matriz del cambio del primer sistema de coordenadas por el segundo y que, por consiguiente (véase el p. 5.1), la matriz A es invertible.

COROLARIO. *El determinante de una matriz cuadrada, formada por elementos de un cuerpo conmutativo, es igual a cero si, y sólo si, las filas de esta matriz son linealmente dependientes.*

Efectivamente, para matrices cuadradas, formadas por elementos de un cuerpo conmutativo, la invertibilidad equivale a la regularidad, es decir, a que el determinante de la matriz sea diferente de cero.

Tomando ahora en consideración la observación hecha al principio de este punto, vemos que un sistema de n vectores de un espacio vectorial de n dimensiones sobre un cuerpo conmutativo forma una base de este espacio si, y sólo si, es diferente de cero el determinante de la matriz formada por las filas de coordenadas de los vectores indicados.

El teorema 1 y su corolario tratan de las filas de una matriz. Sin embargo, estas proposiciones son válidas también para las columnas, siempre que la expresión «relación de dependencia lineal de columnas» sea interpretada como una relación de dependencia lineal respecto a la multiplicación a la derecha de las columnas por los elementos de K. Por ello, siempre que no se diga lo contrario, la multiplicación de filas por elementos de K significará la multiplicación a la izquierda de las filas por los elementos, mientras que la multiplicación de columnas por elementos de K significará la multiplicación a la derecha. Está claro que este convenio sobra si K es un cuerpo conmutativo. Para el caso de cuerpos K no conmutativos el convenio aceptado es de importancia.

Es fácil comprobar que todos los razonamientos que hemos empleado para demostrar el teorema 1 permanecen válidos si la palabra «fila» es sustituida en los mismos por la palabra «columna». Así obtenemos el resultado siguiente.

TEOREMA 1a *Para que una matriz cuadrada, formada por elementos de un cuerpo, sea invertible es necesario y suficiente que sus columnas sean linealmente independientes. El determinante de una matriz cuadrada, formada por elementos de un cuerpo conmutativo, es igual a cero si, y sólo si, las columnas de la matriz son linealmente dependientes.*

Es evidente que si se consideran matrices sobre un cuerpo conmutativo, el teorema la se obtiene del teorema 1 pasando simplemente a la matriz transpuesta.

TEOREMA 2. *Sea* $A = \|\alpha_{ij}\|_{mn}$ *una matriz cualquiera formada por elementos de un cuerpo* K. *Los rangos de la matriz* A *según las filas, las columnas y los menores no varían si* A *se somete a una de las transformaciones siguientes:*

a) *se cambian entre sí cualesquiera filas de la matriz A;*

b) *se cambian entre sí cualesquiera columnas de la matriz A;*

c) *se suman a los elementos de una de las filas de la matriz A los elementos correspondientes de otra fila cualquiera multiplicados a la izquierda por un factor arbitrario fijo* $\lambda \in K$;

d) *se suman a los elementos de una de las columnas de la matriz A los elementos correspondientes de otra columna cualquiera multiplicados a la derecha por un factor arbitrario* $\lambda \in K$.

El rango de la matriz A según las filas es igual al número máximo de elementos linealmente independientes en el conjunto de sus filas $a_1, \ldots, a_m$. Al cambiar el orden de las filas de A sólo alteramos la numeración de las últimas; pero los conceptos de dependencia o independencia lineal de vectores no están relacionados con la numeración. Así mismo, está claro que el rango de la matriz A según las filas no varía al cambiar sus columnas. En efecto, sean $b_1, \ldots, b_m$ las filas de la matriz nueva obtenida después del cambio de columnas. Supongamos que entre las filas de la matriz A existía una relación de dependencia lineal $\lambda_1 a_1 + \ldots + \lambda_m a_m = o$; es evidente que las filas nuevas verificarán la relación análoga $\lambda_1 b_1 + \ldots + \lambda_m b_m = o$ y que, viceversa, la última relación implicará la anterior. Lo mismo ocurrirá si a una columna cualquiera de la matriz A se agrega otra columna multiplicada a la derecha por λ. Finalmente, el número máximo de filas linealmente independientes no cambiará si la i-ésima fila de la matriz A es sustituida por la fila $a_i + \lambda a_j$, ya que los vectores del sistema $a_1, \ldots, a_i + \lambda a_j, \ldots, a_j, \ldots, a_m$ se expresan linealmente en términos de los vectores $a_1, \ldots, a_m$ y los vectores del último sistema se expresan linealmente en términos de los vectores del primer sistema.

Hemos visto que el rango de la matriz A *según las filas* no varía en las transformaciones indicadas en el teorema. Razonamientos análogos demuestran que el rango de la matriz A *según las columnas* no varía en las transformaciones a), b), c) y d). Pasemos ahora a demostrar que tampoco varía el rango de la matriz A *según los menores*, lo cual requerirá una mayor atención.

Supongamos que en la matriz A se cambian entre sí cualesquiera filas o columnas. En este caso los menores de la matriz nueva se obtienen de los menores de la matriz antigua mediante cambios de filas o columnas y sólo debemos comprobar que al cambiar filas o columnas en una matriz invertible de nuevo obtenemos una matriz invertible.

Supongamos, pues, que en una matriz cuadrada $B = \|\beta_{ij}\|_n$ se han cambiado entre sí la primera y la i-ésima filas. El cálculo directo deja constancia de que la matriz nueva puede ser representada en la forma DB, donde

$$D = E_{11} + E_{22} + \ldots + E_{i-1,\,i-1} + E_{i1} + E_{i+1,\,i+1} + \ldots + E_{nn}.$$

Aquí E_{ij} es la matriz cuyo elemento de la i-ésima fila y j-ésima columna es la unidad, mientras que todos los demás elementos suyos son ceros. Mediante cálculo directo podemos comprobar que

$$DD=E \quad \text{y} \quad D^{-1}=D.$$

Por esto, si la matriz B tiene inversa, la matriz $B^{-1}D$ será la inversa de DB. Análogamente, la matriz BD se obtiene de B cambiando entre sí la primera y la i-ésima columnas. Si B es invertible, la matriz BD también tiene la inversa, a saber, DB^{-1}.

Hemos demostrado que el rango de menores no varía al cambiar entre sí filas y columnas. Supongamos que el rango de menores de B es igual a r y que a la i-ésima fila de B se ha agregado la j-ésima fila multiplicada por λ. Por hipótesis, la matriz B posee un menor invertible de orden r perteneciente a r filas y a r columnas determinadas. Para no complicar sin necesidad la notación, aceptemos que este menor pertenece a las r filas primeras y a las r columnas primeras de la matriz B. Si la i-ésima fila que se altera no pertenece a las r primeras filas, el menor indicado de orden r continuará siendo un menor de la matriz nueva. Por esto el rango de la matriz nueva es $\geqslant r$. Supongamos que la i-ésima fila es una de las r primeras filas. Consideremos los elementos de las matrices nueva y antigua que se encuentran en las r primeras columnas, en las r primeras filas y en la j-ésima fila ($j>r$). Obtendremos las matrices

$$P=\begin{bmatrix} \beta_{11} & \dots & \beta_{1r} \\ \dots & \dots & \dots \\ \beta_{i1} & \dots & \beta_{ir} \\ \dots & \dots & \dots \\ \beta_{r1} & \dots & \beta_{rr} \\ \beta_{j1} & \dots & \beta_{jr} \end{bmatrix} \quad \text{y} \quad Q=\begin{bmatrix} \beta_{11} & \dots & \beta_{1r} \\ \dots & \dots & \dots \\ \beta_{i1}+\lambda\beta_{j1} & \dots & \beta_{ir}+\lambda\beta_{jr} \\ \dots & \dots & \dots \\ \beta_{r1} & \dots & \beta_{rr} \\ \beta_{j1} & \dots & \beta_{jr} \end{bmatrix}.$$

En virtud del teorema 1, las primeras r filas de la matriz P son linealmente independientes y, por ello, el rango de P según las filas es igual a r. De acuerdo con lo demostrado, de aquí se deduce que el rango de la matriz Q según las filas es también igual a r. Las primera, ..., $(i-1)$-ésima, $(i+1)$-ésima, ..., r-ésima filas de la matriz Q son desde luego linealmente independientes, ya que forman parte del sistema linealmente independiente de las r primeras filas de la matriz Q. Por esto, o bien las r primeras filas de la matriz Q son linealmente independientes o bien la i-ésima fila se expresa linealmente en términos de las restantes y, por consiguiente, son linealmente independientes las restantes r filas de la matriz Q. En

otras palabras, uno de los menores

$$\begin{bmatrix} \beta_{11} & \cdots & \beta_{1r} \\ \cdots & \cdots & \cdots \\ \beta_{i1}+\lambda\beta_{j1} & \cdots & \beta_{ir}+\lambda\beta_{jr} \\ \cdots & \cdots & \cdots \\ \beta_{r1} & \cdots & \beta_{rr} \end{bmatrix} \quad \text{o} \quad \begin{bmatrix} \beta_{11} & \cdots & \beta_{1r} \\ \cdots & \cdots & \cdots \\ \beta_{i-1,1} & \cdots & \beta_{i-1,r} \\ \beta_{i+1,1} & \cdots & \beta_{i+1,r} \\ \cdots & \cdots & \cdots \\ \beta_{r1} & \cdots & \beta_{rr} \\ \beta_{j1} & \cdots & \beta_{jr} \end{bmatrix}$$

de la matriz Q está formado por filas linealmente independientes. En virtud del teorema 1, este menor es invertible y el rango de menores de la matriz Q es no menor de r.

Hemos demostrado que la transformación de tipo c) no disminuye el rango de menores de una matriz. Pero, si la matriz C se obtiene de la matriz A agregando a su i-ésima fila la j-ésima fila multiplicada por λ, la matriz A se obtiene de C sumando a la i-ésima fila de la matriz C su j-ésima fila multiplicada por $-\lambda$. En una y otra transformación los rangos de menores no disminuyen y, por consiguiente, no varían. Resta considerar el caso de la transformación d); pero para ella la invariabilidad del rango de la matriz A se demuestra de la misma forma que la invariabilidad en la transformación c).

Del teorema 2 mediante razonamientos sencillos se deduce el siguiente teorema principal.

TEOREMA 3 (SOBRE EL RANGO DE UNA MATRIZ). *Para una matriz arbitraria* $A = \|\alpha_{ij}\|_{mn}$, *formada por elementos de un cuerpo, el rango según las filas, el rango según las columnas y el rango según los menores coinciden.*

La proposición es evidente si todos los elementos de la matriz A son iguales a cero. Por esto, supondremos que A contiene un elemento α_{ij} diferente de cero. Cambiando entre sí la primera y la i-ésima filas y la primera y la j-ésima columnas de la matriz A, obtendremos una matriz $B = \|\beta_{ij}\|$ en la que $\beta_{11} = \alpha_{ij} \neq 0$ y, además, en virtud del teorema 2, los tres rangos de la matriz B serán iguales a los rangos correspondientes de la matriz A. Sumando ahora a la segunda, ..., m-ésima filas de la matriz B su primera fila multiplicada a la izquierda por $-\beta_{21}\beta_{11}^{-1}$, ..., $-\beta_{m1}\beta_{11}^{-1}$, respectivamente, obtendremos una matriz de tipo

$$C_1 = \begin{vmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1n} \\ 0 & \gamma_{22} & \cdots & \gamma_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \gamma_{m2} & \cdots & \gamma_{mn} \end{vmatrix}.$$

con la particularidad de que los tres rangos de la matriz C_1 serán los mismos que los de la matriz A. Si todos los γ_{ij} son iguales

a cero, no realizamos ninguna transformación más. En cambio, si $\gamma_{ij} \neq 0$, $(i \geqslant 2,\ j \geqslant 2)$, cambiamos entre sí la segunda y la i-ésima filas y la segunda y la j-ésima columnas de la matriz C_1, obteniendo así una matriz en la que el elemento γ_{22} sea diferente de cero. Sumando entonces a la tercera, ..., m-ésima filas la segunda fila multiplicada por $-\gamma_{32}\gamma_{22}^{-1}$, ..., $-\gamma_{m2}\gamma_{22}^{-1}$, respectivamente, obtendremos una matriz de tipo

$$C_2 = \begin{bmatrix} \delta_{11} & \delta_{12} & \delta_{13} & \dots & \delta_{1n} \\ 0 & \delta_{22} & \delta_{23} & \dots & \delta_{2n} \\ 0 & 0 & \delta_{33} & \dots & \delta_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \delta_{m3} & \dots & \delta_{mn} \end{bmatrix} \quad (\delta_{11}\delta_{22} \neq 0).$$

Los rangos de la matriz C_2 también coincidirán con los rangos correspondientes de la matriz A. Continuando este proceso, al cabo de un número de pasos no mayor que m, obtendremos una matriz de tipo

$$C_r = \begin{bmatrix} \mu_{11} & \mu_{12} & \dots & \mu_{1r} & \mu_{1,r+1} & \dots & \mu_{1n} \\ 0 & \mu_{22} & \dots & \mu_{2r} & \mu_{2,r+1} & \dots & \mu_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \mu_{rr} & \mu_{r,r+1} & \dots & \mu_{rn} \\ 0 & 0 & \dots & 0\ 0 & & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0\ 0 & & \dots & 0 \end{bmatrix} \quad (\mu_{11} \dots \mu_{rr} \neq 0).$$

cuyos rangos coincidirán también con los rangos correspondientes de la matriz A. Sin embargo, de la forma de la matriz C_r se deduce directamente que los tres rangos de esta matriz son iguales a un mismo número r. Por esto, los tres rangos de la matriz A también tienen un mismo valor r.

En el p. 2.3 se ha demostrado que una matriz cuadrada con elementos de un cuerpo conmutativo es invertible si, y sólo si, tiene el determinante diferente de cero. Por esto, para matrices sobre un cuerpo conmutativo el teorema 3 se puede enunciar del modo siguiente:

TEOREMA 3a. *En toda matriz con elementos de un cuerpo conmutativo el número máximo de filas linealmente independientes es igual al número máximo de columnas linealmente independientes e igual al orden máximo de sus menores de determinante diferente de cero.*

La determinación práctica del rango de una matriz se realiza generalmente aplicando el método indicado en la demostración del teorema 3 o alguna de sus modificaciones convenientes.

Con frecuencia resultan útiles las dos observaciones siguientes que se deducen directamente del teorema 3.

TEOREMA 4. *Cualesquiera que sean dos matrices* $A=\|\alpha_{ij}\|_{mn}$ *y* $B=\|\beta_{jk}\|_{np}$ *con elementos de un cuerpo arbitrario el rango del producto* AB *es no mayor que el rango de la matriz* A *y no mayor que el rango de la matriz* B.

Efectivamente, de la definición del producto se ve que las filas de la matriz AB son combinaciones lineales de las filas de la matriz B y, por consiguiente, el número máximo de filas linealmente independientes de la matriz AB no puede pasar del número máximo de filas linealmente independientes de la matriz B. Análogamente, las columnas de la matriz AB son combinaciones lineales de las columnas de la matriz A y, por consiguiente, el rango de la matriz AB según las columnas es no mayor que el rango según las columnas de la matriz A.

TEOREMA 5 *Si los elementos de la matriz* A *pertenecen a un cuerpo conmutativo, su rango coincide con el rango de la matriz transpuesta* A'.

Todos los menores de la matriz A' se obtienen por transposición de los menores de la matriz A. Puesto que los elementos de los menores se toman en un cuerpo conmutativo, de la transposición de un menor invertible resulta un menor invertible y, por consiguiente, el rango (según los menores) de la matriz A' coincide con el rango (según los menores) de la matriz A.

En los ejercicios de la pág. 101 se da un ejemplo de una matriz invertible, formada por elementos de un cuerpo no conmutativo, cuya matriz transpuesta no es invertible.

5.3. Sistemas generales de ecuaciones lineales. Consideremos un cuerpo K y un sistema arbitrario de ecuaciones de tipo

$$\left.\begin{aligned}
&\alpha_{11}\xi_1+\alpha_{12}\xi_2+\ldots+\alpha_{1n}\xi_n=\beta_1,\\
&\alpha_{21}\xi_1+\alpha_{22}\xi_2+\ldots+\alpha_{2n}\xi_n=\beta_2,\\
&\qquad\ldots\ldots\ldots\ldots\ldots\\
&\alpha_{m1}\xi_1+\alpha_{m2}\xi_2+\ldots+\alpha_{mn}\xi_n=\beta_m,
\end{aligned}\right\} \tag{1}$$

donde α_{ij} y β_i son elementos dados de K y $\xi_1, \xi_2, \ldots, \xi_n$ son incógnitas, cuyo número n puede ser mayor, menor o igual al número m de ecuaciones. En el p. 2.4 ha sido expuesto uno de los algoritmos de la resolución de estos sistemas, el método de eliminación de incógnitas. El teorema sobre el rango de una matriz, demostrado en el punto anterior, ofrece un acceso diferente al estudio de los sistemas de ecuaciones lineales.

Los coeficientes α_{ij} y los términos independientes β_i de las ecuaciones (1) se pueden disponer de un modo natural formando

dos matrices:

$$A=\begin{bmatrix}\alpha_{11} & \alpha_{12} & \dots & \alpha_{1n}\\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2n}\\ \dots & \dots & \dots & \dots\\ \alpha_{m1} & \alpha_{m2} & \dots & \alpha_{mn}\end{bmatrix} \quad \text{y} \quad B=\begin{bmatrix}\alpha_{11} & \alpha_{12} & \dots & \alpha_{1n} & \beta_1\\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2n} & \beta_2\\ \dots & \dots & \dots & \dots & \dots\\ \alpha_{m1} & \alpha_{m2} & \dots & \alpha_{mn} & \beta_m\end{bmatrix},$$

llamadas *matriz principal* y *matriz ampliada* del sistema (1). A toda ecuación del sistema (1) corresponde una fila determinada en las matrices A y B. El intercambio de ecuaciones en el sistema (1) lleva al intercambio correspondiente de filas en las matrices A y B y la modificación de la numeración de las incógnitas lleva al intercambio de columnas en las matrices indicadas.

Se dice que la i-ésima ecuación del sistema (1) depende linealmente de las i_1-ésima, ..., i_s-ésima ecuaciones de este sistema, si la i-ésima fila de la matriz B es una combinación lineal de sus i_1-ésima, ..., i_s-ésima filas. Puesto que el cuerpo K no se supone conmutativo, las filas de las matrices se multiplican por elementos de K siempre a la izquierda, mientras que las columnas, a la derecha (véase el p. 5.2).

Se llama *solución* del sistema de ecuaciones (1) una sucesión $\xi_1^0, \dots, \xi_n^0$ de elementos del cuerpo K que al ser introducidos en las ecuaciones (1) en lugar de las letras $\xi_1, \dots, \xi_n$ hacen válidas todas las igualdades. Si disponemos los elementos $\xi_1, \dots, \xi_n$ en una columna $[\xi_1, \dots, \xi_n]' = x$, el sistema (1) puede ser representado en la forma matricial

$$A \cdot x = b, \tag{2}$$

donde b es la columna $[\beta_1, \dots, \beta_n]'$ de los términos independientes.

A) *Si la i-ésima fila de la matriz B es una combinación lineal de sus restantes filas, entonces suprimiendo en el sistema* (1) *la i-ésima ecuación obtendremos un sistema reducido de ecuaciones que tiene el mismo conjunto de soluciones que el sistema* (1).

Está claro que toda solución del sistema (1) es también una solución del sistema reducido. Recíprocamente, supongamos que la i-ésima fila de la matriz B es igual a la suma de las i_1-ésima, ..., i_s-ésima filas multiplicadas a la izquierda, respectivamente, por los elementos $\lambda_1, \dots, \lambda_s$ de K ($i \neq i_1, \dots, i_s$) y supongamos que los valores $\xi_1^0, \dots, \xi_n^0$ satisfacen las i_1-ésima, ..., i_s-ésima ecuaciones de (1). Entonces, multiplicando estas ecuaciones a la izquierda por $\lambda_1, \dots \lambda_s$, respectivamente, y sumándolas término por término, obtendremos la i-ésima ecuación.

B) *Si el rango (según las filas) de la matriz B es igual a r, el sistema de ecuaciones* (1) *contiene un subsistema de r ecuaciones linealmente independientes que posee el mismo conjunto de soluciones que el sistema* (1).

Puesto que el rango de la matriz B es igual a r, la matriz B posee r filas linealmente independientes y todas sus filas restantes se expresan linealmente en términos de éstas. Tomando las ecuaciones que corresponden a las r filas linealmente independientes señaladas, obtendremos el subsistema deseado.

C) *Si el sistema de ecuaciones* (1) *tiene una solución* $\xi_1^0, \ldots, \xi_n^0$, *el rango de la matriz ampliada* B *coincide con el rango de la matriz principal* A.

Si las igualdades (1) son válidas para $\xi_1^0, \ldots, \xi_n^0$, esto demuestra que la última columna de la matriz B es igual a la suma de sus primera, ..., n-ésima columnas multiplicadas a la derecha por $\xi_1^0, \ldots, \xi_n^0$, respectivamente. Por esto, el número máximo de columnas linealmente independientes de la matriz B es igual al número máximo de columnas linealmente independientes de la matriz A, es decir, el rango según las columnas de la matriz B es igual al rango según las columnas de la matriz A. Las palabras «según las columnas» pueden ser omitidas, ya que en el p. 5.2 se ha demostrado que el rango según las columnas coincide con el rango según las filas.

De las proposiciones B) y C) resulta fácilmente el siguiente teorema principal.

TEOREMA DE KRONECKER–CAPELLI.[1]) *Para que el sistema de ecuaciones* (1) *tenga solución, es necesario y suficiente que el rango de la matriz ampliada sea igual al rango de la matriz principal de este sistema. Si los rangos de las matrices principal y ampliada coinciden con el número de incógnitas, el sistema tiene una solución única. Si el rango* r *de las matrices principal y ampliada es inferior al número* n *de incógnitas, el sistema* (1) *tiene más de una solución y es equivalente a un sistema de tipo*

$$\left.\begin{aligned} \xi_{i_1} &= \gamma_{11}\xi_{j_1} + \ldots + \gamma_{1,\,n-r}\xi_{j_{n-r}} + \gamma_{1,} \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ \xi_{i_r} &= \gamma_{r1}\xi_{j_1} + \ldots + \gamma_{r,\,n-r}\xi_{j_{n-r}} + \gamma_{r,} \end{aligned}\right\} \qquad (3)$$

donde $(i_1, \ldots, i_r, j_1, \ldots, j_{n-r})$ *es una permutación adecuada de los números* $1, 2, \ldots, n$.

En otras palabras, en el último caso entre las incógnitas $\xi_1, \ldots, \xi_n$ se pueden escoger $n - r$ incógnitas $\xi_{j_1}, \ldots, \xi_{j_{n-r}}$, llamadas libres, y dándoles valores arbitrarios de K se pueden encontrar para las restantes incógnitas unos valores únicos que satisfagan el sistema (1).

La necesidad de las condiciones se deduce de la proposición C). Supongamos, por esto, que los rangos de las matrices A y B coinciden y son iguales a r. La matriz A contiene un total de n columnas, de modo que $r \leqslant n$. Por hipótesis, la matriz A posee un

[1]) En algunos libros españoles este teorema se conoce como el teorema de Rouché—Frobenius. *(N. del Tr.)*

menor invertible M de orden r (véase el p. 5.2) que pertenece a las k_1-ésima, ..., k_r-ésima filas y a las i_1-ésima, ..., i_r-ésima columnas de la matriz A. De aquí se deduce que la k_1-ésima, ..., la k_r-ésima filas de la matriz B son linealmente independientes, mientras que sus filas restantes son combinaciones lineales de las filas señaladas. Por esto, dejando solamente la k_1-ésima, ..., la k_r-ésima ecuaciones del sistema (1) y suprimiendo todas las demás ecuaciones, obtendremos un sistema reducido que tendrá las mismas soluciones que el sistema (1). Dejando ahora en los primeros miembros de cada una de las ecuaciones del sistema reducido los términos que contienen las incógnitas $\xi_{i_1}, \ldots, \xi_{i_r}$ y pasando al otro miembro todos los demás términos, llevaremos el sistema reducido a la forma

$$M \cdot \begin{bmatrix} \xi_{i_1} \\ \cdot \\ \cdot \\ \cdot \\ \xi_{i_r} \end{bmatrix} = \begin{bmatrix} c_1 \\ \cdot \\ \cdot \\ \cdot \\ c_r \end{bmatrix}, \tag{4}$$

donde

$$c_s = -\alpha_{k_s, i_1}\xi_{i_1} - \ldots - \alpha_{k_s, i_{n-r}}\xi_{i_{n-r}} + \beta_{k_s} \quad (s = 1, \ldots, r).$$

Por hipótesis, el menor M es una matriz invertible. Multiplicando ambos miembros de la igualdad (4) a la izquierda por M^{-1}, obtenemos

$$\begin{bmatrix} \xi_{i_1} \\ \cdot \\ \cdot \\ \cdot \\ \xi_{i_r} \end{bmatrix} = M^{-1} \begin{bmatrix} c_1 \\ \cdot \\ \cdot \\ \cdot \\ c_r \end{bmatrix}.$$

Realizando la multiplicación en el segundo miembro, llegamos al sistema de tipo (3). El teorema queda demostrado.

El sistema de ecuaciones (1) suele llamarse *determinado*, *indeterminado* y *contradictorio* según tenga, respectivamente, solución única, más de una solución y no tenga solución. Notemos que de acuerdo a esta terminología un sistema, que no es determinado, es o bien indeterminado o bien contradictorio.

COROLARIO 1. *Si el número de ecuaciones del sistema* (1) *es inferior al número de incógnitas, el sistema* (1) *es o bien indeterminado o bien contradictorio.*

COROLARIO 2. *Un sistema de n ecuaciones lineales con n incógnitas es determinado si, y sólo si, la matriz principal de este sistema es invertible.*

Ambos corolarios se desprenden directamente del teorema de Kronecker—Capelli.

Hemos considerado el sistema de ecuaciones lineales (1) en el que los coeficientes aparecen a la izquierda de las incógnitas. ¿Qué puede decirse acerca del sistema de ecuaciones

$$\left.\begin{aligned} \xi_1\alpha_{11}+\xi_2\alpha_{21}+\ldots+\xi_n\alpha_{n1}&=\beta_1,\\ \ldots\ldots\ldots\ldots\ldots\ldots&\\ \xi_1\alpha_{1m}+\xi_2\alpha_{2m}+\ldots+\xi_n\alpha_{nm}&=\beta_m \end{aligned}\right\} \qquad (5)$$

en el que los coeficientes figuran a la derecha de las incógnitas? Este sistema puede ser representado en la forma matricial

$$[\xi_1, \ldots, \xi_n]\cdot A=[\beta_1, \ldots, \beta_m],$$

donde $A=\|\alpha_{ij}\|$ es la matriz de los coeficientes de las incógnitas en el sistema (5). Prestemos atención a que en la matriz A los coeficientes de la i-ésima ecuación forman la i-ésima columna y no la i-ésima fila como sucedía anteriormente. Todos los razonamientos, que hemos realizado al demostrar el teorema de Kronecker—Capelli, siguen siendo válidos también para el sistema (5), siempre que se cambien entre sí en ellos las palabras «columna» «y «fila» y siempre que por la matriz ampliada del sistema (5) se comprenda la matriz

$$B=\begin{bmatrix} \alpha_{11} & \ldots & \alpha_{1m} \\ \ldots & \ldots & \ldots \\ \alpha_{n1} & \ldots & \alpha_{nm} \\ \beta_1 & \ldots & \beta_m \end{bmatrix}.$$

Por consiguiente, para que el sistema (5) sea compatible es necesario y suficiente que el rango de la matriz B coincida con el rango de la matriz A. Si los rangos de las matrices A y B coinciden y son iguales al número de incógnitas, el sistema (5) tiene solución única. Si los rangos de las matrices A y B coinciden y son iguales a un número r menor que el número n de las incógnitas, entre las incógnitas podrán encontrarse $n-r$ incógnitas libres $\xi_{i_1}, \ldots, \xi_{i_{n-r}}$ y todas las demás incógnitas $\xi_{i_1}, \ldots, \xi_{i_r}$ se expresarán en términos de éstas por fórmulas de tipo

$$\xi_{i_s}=\xi_{i_1}\gamma_{1s}+\ldots+\xi_{i_{n-r}}\gamma_{n-r,s}+\gamma_s \quad (s=1, \ldots, r).$$

Está claro que toda diferencia entre los sistemas (1) y (5) desaparece, si se consideran ecuaciones lineales sobre un cuerpo conmutativo. En el caso de cuerpos no conmutativos K (por ejemplo, para los cuaternios (véase el p. 1.5)), además de los sistemas standard de tipo (1) y (5) se consideran también ecuaciones de tipo «mixto» como es, por ejemplo, la ecuación $\alpha\xi-\xi\alpha=0$. Sin embargo, los problemas relacionados con la resolución de estas ecuaciones tienen un carácter específico y se salen de los márgenes del Algebra lineal propiamente dicha.

Complementos y ejercicios

1. En el espacio de filas de longitud tres sobre el cuerpo de los números racionales se toma un sistema de coordenadas formado por las filas [1, 3, 5], [6, 3, 2] y [3, 1, 0]. ¿Qué filas de coordenadas tienen en este sistema los vectores [3, 7, 1], [0, 0, 1], [2, 3, 5] y [1, 1, 1]?

2. El espacio $\mathfrak{L}$ está formado por los polinomios en λ de grado no mayor que n. Demuéstrese que los polinomios 1, $\lambda-1$, $(\lambda-1)^2$, ..., $(\lambda-1)^n$ constituyen una base de $\mathfrak{L}$. Hállense en esta base las filas de coordenadas de los polinomios $2-3\lambda+\lambda^2$ y λ^n.

3. En el plano se toma un sistema de coordenadas, compuesto por dos vectores mutuamente perpendiculares de longitud 1, y después se realiza una transformación de coordenadas de matriz $\begin{bmatrix}\alpha & \beta\\ \gamma & \delta\end{bmatrix}$. ¿Qué condiciones deben verificar los números α, β, γ y δ para que los nuevos vectores coordenados sean mutuamente perpendiculares y de longitud 1?

4. Hállense los rangos de las matrices

$$\begin{bmatrix} 1 & 0 & 2 & 5 & 7\\ -2 & 1 & 1 & 2 & 3\\ 2 & 1 & 7 & 17 & 24\\ 1 & 3 & 15 & 41 & 55\end{bmatrix}\cdot\begin{bmatrix} 1 & 1 & 1 & 1\\ x & y & z & u\\ x^2 & y^2 & z^2 & u^2\\ x^3 & y^3 & z^3 & u^3\end{bmatrix}\text{ y }\begin{bmatrix} 1+\lambda & 1 & 1 & 1 & 1\\ 1 & 1+\lambda & 1 & 1 & 1\\ 1 & 1 & 1+\lambda & 1 & 1\\ 1 & 1 & 1 & 1+\lambda & 1\\ 1 & 1 & 1 & 1 & 1+\lambda\end{bmatrix}.$$

5. Demuéstrese que en el cuerpo de los cuaternios K (p.1.4) el sistema de ecuaciones lineales «a la derecha»

$$\left.\begin{aligned} i\xi_1+j\xi_2&=1,\\ k\xi_1-\xi_2&=i\end{aligned}\right\}$$

tiene solución única, mientras que el sistema de ecuaciones «a la izquierda»

$$\left.\begin{aligned} \xi_1 i+\xi_2 j&=1,\\ \xi_1 k-\xi_2&=i\end{aligned}\right\}$$

no tiene soluciones. Tenemos así un ejemplo de una matriz invertible sobre un cuerpo, cuya matriz transpuesta no es invertible.

6. Hasta el momento entendíamos la independencia lineal de filas como la independencia lineal respecto a la multiplicación de filas a la izquierda por los elementos del cuerpo K y la independencia lineal de las columnas como la independencia respecto a la multiplicación de columnas a la derecha por los elementos del cuerpo. Esto está relacionado con la regla inicial de multiplicación de las matrices: las filas de la primera se multiplican por las columnas de la segunda. Por supuesto, la teoría no cambiará si las matrices se multiplican por la regla «izquierda»: las columnas de la primera se multiplican por las filas de la segunda. Pero en este caso, habrá que considerar la dependencia lineal de las filas respecto a la multiplicación a la derecha por los elementos de K y la dependencia lineal de las columnas respecto a la multiplicación a la izquierda por los elementos de K. Como resultado, obtendremos los rangos de una matriz según las filas a la derecha y según las columnas a la izquierda. Ambos rangos coincidirán, pero en el caso general serán diferentes de los rangos corrientes según las filas (a la izquierda) y según las columnas (a la derecha). Como ejemplo puede servir la matriz de cuaternios

$$\begin{bmatrix} i & j\\ k & -1\end{bmatrix}$$

del ejercicio 5, en la que el rango a la izquierda según las filas es igual a 2, mientras que el rango a la derecha según las filas es igual a 1.

7. Sea A una matriz cuadrada invertible de orden n y sean B y C matrices arbitrarias compuestas, respectivamente, por n filas y n columnas. En estas con-

diciones, se tiene

$$\text{rango}\,(AB) = \text{rango}\,B,$$
$$\text{rango}\,(CA) = \text{rango}\,C.$$

8. Dése un ejemplo de matrices cuadradas A y B de un mismo orden 2, para las cuales se tenga $\text{rango}\,(AB) \neq \text{rango}\,(BA)$.

9. Demuéstrese que para cualesquiera matrices cuadradas A y B de orden n se tiene

$$\text{rango}\,(AB) \geqslant \text{rango}\,A + \text{rango}\,B - n.$$

§ 6. Subespacios lineales

6.1. Intersección y suma de subespacios. Un conjunto no vacío $\mathfrak{A}$ de vectores de un espacio lineal $\mathfrak{L}$ se llama *subespacio lineal* de este espacio, si se cumplen las dos condiciones siguientes:

1° *si* $\mathfrak{A}$ *contiene un vector* a, *también contiene todos los múltiplos* λa, *donde* λ *es un número del cuerpo de coeficientes*;

2° *si* $\mathfrak{A}$ *contiene unos vectores* a *y* b, *también contiene su suma* $a + b$.

Está claro que ambas condiciones equivalen a la siguiente: *si* $\mathfrak{A}$ *contiene unos vectores* a *y* b, *también contiene cualquier combinación lineal* $\lambda a + \mu b$ *de los mismos*.

De estas definiciones se desprende que todo subespacio lineal $\mathfrak{A}$ contiene el vector nulo y todas las combinaciones lineales de cualesquiera vectores suyos.

En el p. 4.3 se ha señalado que todo espacio lineal sobre un cuerpo K es un álgebra en la que las operaciones principales son: la adición de vectores, la inversión de vectores y la multiplicación de los vectores (a la izquierda) por elementos de K. Puesto que la inversión de un vector equivale a su multiplicación por el elemento -1 de K, las condiciones 1° y 2° significan simplemente que *los subespacios lineales del espacio* $\mathfrak{L}$ *son subálgebras del álgebra* $\mathfrak{L}$.

El conjunto compuesto solamente del vector nulo posee las propiedades 1° y 2° y, por consiguiente, es un subespacio lineal de $\mathfrak{L}$. Este subespacio se llama subespacio *nulo*. Por otro lado, el propio espacio $\mathfrak{L}$ puede ser considerado como un subespacio lineal de sí mismo. El subespacio nulo y $\mathfrak{L}$ suelen llamarse subespacios *triviales* del espacio $\mathfrak{L}$. Todos los demás subespacios se denominan *no triviales* o *propios*.

El método más sencillo de obtener subespacios lineales consiste en lo siguiente. En el espacio lineal dado $\mathfrak{L}$ se toman unos vectores arbitrarios $a_1, a_2, \ldots, a_m$ y se consideran todas las combinaciones lineales de los mismos

$$\alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_m a_m.$$

Sea $\mathfrak{A}$ el conjunto de estas combinaciones. Puesto que la suma de

combinaciones lineales de $a_1, \ldots, a_m$ y el producto de una combinación lineal por un número son combinaciones lineales de $a_1, \ldots, a_m$, tenemos que $\mathfrak{A}$ es un subespacio lineal. Los vectores $a_1, \ldots, a_m$ son los generadores de este subespacio. Suprimiendo aquellos que dependen linealmente de los anteriores, obtendremos un sistema linealmente independiente de generadores del subespacio $\mathfrak{A}$, es decir, una base de $\mathfrak{A}$. Pero el número de vectores de una base coincide con la dimensión del espacio y, por ello, *la dimensión del subespacio* $\mathfrak{A}$ *es igual al número máximo de vectores linealmente independientes que contiene el sistema* $a_1, \ldots, a_m$. A veces se dice que el subespacio $\mathfrak{A}$ es el subespacio *tendido* sobre los vectores $a_1, \ldots, a_m$. Por consiguiente, la dimensión del subespacio tendido sobre el sistema de vectores $a_1, \ldots, a_m$ es igual al número máximo de vectores linealmente independientes que contiene este sistema.

El método de tender los subespacios sobre un sistema de vectores dado es general: *todo subespacio lineal* $\mathfrak{A}$ *de un espacio lineal* $\mathfrak{L}$ *es el subespacio tendido sobre su base.*

Como que el número de vectores linealmente independientes de $\mathfrak{L}$ no puede superar la dimensión de $\mathfrak{L}$, de aquí se ve que la dimensión de un subespacio lineal no puede superar la dimensión del espacio que lo envuelve. Es más, si la dimensión del espacio lineal $\mathfrak{L}$ es igual a la dimensión del subespacio $\mathfrak{A}$, toda base de $\mathfrak{A}$ será también una base de $\mathfrak{L}$. Por esto, todo vector de $\mathfrak{L}$ se expresará linealmente en términos de una base del subespacio $\mathfrak{A}$, es decir, $\mathfrak{A}$ coincidirá con $\mathfrak{L}$. Por consiguiente, la dimensión de todo subespacio lineal propio es inferior a la dimensión del espacio que lo envuelve.

A título de ejemplo consideremos el espacio corriente $\mathfrak{R}$, formado por los segmentos orientados que parten de un punto O. La dimensión del espacio $\mathfrak{R}$ es igual a tres y, por esto, los subespacios propios pueden ser de dimensión uno o dos. Los subespacios de dimensión uno deben ser tendidos sobre un vector no nulo a, es decir, deben ser el conjunto de los multiplos αa del segmento a. Pero todos los segmentos de tipo αa se hallan sobre la recta que contiene al vector a. Por consiguiente, *los subespacios de* $\mathfrak{R}$ *de dimensión una son las rectas que pasan por el punto* O.

Los subespacios de dos dimensiones deben ser tendidos sobre dos vectores linealmente independientes, es decir, sobre dos vectores a y b que no pertenecen a una misma recta. Sea $\mathfrak{A}$ el plano que pasa por los vectores a y b. Entonces, todas las combinaciones lineales $\alpha a + \beta b$ pertenecerán al plano $\mathfrak{A}$ y, por otro lado, todo vector perteneciente a $\mathfrak{A}$ será una combinación lineal de los vectores a y b. Por consiguiente, el subespacio tendido sobre los vectores a y b será el conjunto de vectores pertenecientes al plano $\mathfrak{A}$. Luego, *los subespacios de dos dimensiones del espacio* $\mathfrak{R}$ *son los planos que pasan por el punto* O.

Los espacios de dimensión mayor que tres no admiten una interpretación geométrica tan clara. Sin embargo, a ellos también se aplica la terminología geométrica llamando *rectas* a los subespacios lineales de una dimensión, *planos* a los subespacios de dos dimensiones y *planos de k dimensiones* a los subespacios de dimensión k, para $k \geqslant 3$. Los subespacios lineales de dimensión menor en 1 que la dimensión del espacio llevan el nombre especial de *hiperplanos*.

Con los subespacios de un espacio lineal dado se pueden efectuar determinadas operaciones; las más importantes de éstas son la adición y la intersección. Se llama *intersección* de los subespacios $\mathfrak{A}$, $\mathfrak{B}$, ... de un espacio $\mathfrak{L}$ el conjunto $\mathfrak{P}$ de los vectores que pertenecen simultáneamente a todos estos espacios. La operación de intersección se indica por el símbolo $\cap$, de modo que

$$\mathfrak{P} = \mathfrak{A} \cap \mathfrak{B} \cap \ldots$$

La intersección de cualquier número de subespacios lineales de un espacio $\mathfrak{L}$ es un subespacio lineal de este espacio.

En efecto, el vector nulo pertenece a cada uno de los subespacios dados $\mathfrak{A}$, $\mathfrak{B}$, ... Por esto, pertenece también a la intersección $\mathfrak{P}$ de los mismos que es, por consiguiente, un conjunto no vacío. Por otro lado, si unos vectores a y b pertenecen a la intersección $\mathfrak{P}$, estos vectores, y con ellos cualquier combinación lineal $\alpha a + \beta b$ de los mismos, pertenecerán a cada uno de los subespacios $\mathfrak{A}$, $\mathfrak{B}$, ... Por consiguiente, $\alpha a + \beta b$ está contenido en $\mathfrak{P}$, es decir, $\mathfrak{P}$ es un subespacio lineal.

Se llama *suma* de un número finito de subespacios lineales $\mathfrak{A}_1$, $\mathfrak{A}_2$, ..., $\mathfrak{A}_s$ del espacio $\mathfrak{L}$ el conjunto de vectores que pueden ser representados en la forma

$$a = a_1 + a_2 + \ldots + a_s, \tag{1}$$

donde a_i es un vector de $\mathfrak{A}_i$ ($i = 1, \ldots, s$). La operación de adición de subespacios se indica por el símbolo $+$.

La suma de un número finito de subespacios lineales de un espacio $\mathfrak{L}$ es también un subespacio lineal; contiene todos los vectores de los subespacios dados, así como todas sus combinaciones lineales.

En efecto, sean $\mathfrak{A}_1, \ldots, \mathfrak{A}_s$ los subespacios lineales dados y sea $\mathfrak{A} = \mathfrak{A}_1 + \ldots + \mathfrak{A}_s$. Si a y b son unos vectores de $\mathfrak{A}$, ello significa que pueden ser representados en la forma

$$a = a_1 + a_2 + \ldots + a_s \quad \text{y} \quad b = b_1 + b_2 + \ldots + b_s,$$

donde a_i y b_i pertenecen a $\mathfrak{A}_i$ ($i = 1, \ldots, s$). Pero, en este caso, para cualesquiera α y β de K la expresión

$$\alpha a + \beta b = (\alpha a_1 + \beta b_1) + (\alpha a_2 + \beta b_2) + \ldots + (\alpha a_s + \beta b_s)$$

es la descomposición en la forma (1) del vector $\alpha a + \beta b$, ya que la suma $\alpha a_i + \beta b_i$ pertenece a $\mathfrak{A}_i$. Por consiguiente, $\alpha a + \beta b$ pertenece

a $\mathfrak{A}$ y $\mathfrak{A}$ es un subespacio lineal. Tomando ahora en (1) $a_j = o$ para $j \neq i$, vemos que a_i estará contenido en $\mathfrak{A}$, es decir, $\mathfrak{A}$ contiene todos los vectores de $\mathfrak{A}_i$.

Indiquemos, finalmente, sin demostración las propiedades siguientes de la suma de subespacios, que se desprenden directamente de su definición:

1° $\mathfrak{A} + \mathfrak{B} = \mathfrak{B} + \mathfrak{A}$;

2° $\mathfrak{A} + (\mathfrak{B} + \mathfrak{C}) = (\mathfrak{A} + \mathfrak{B}) + \mathfrak{C}$;

3° *si* $\mathfrak{A}$ *está contenido en un subespacio* $\mathfrak{B}$, *se tiene* $\mathfrak{A} + \mathfrak{B} = \mathfrak{B}$.

Considerando la suma de dos subespacios lineales arbitrarios $\mathfrak{A}$ y $\mathfrak{B}$, podremos ver fácilmente que su dimensión depende no sólo de la dimensión de los subespacios $\mathfrak{A}$ y $\mathfrak{B}$, sino también de cuán grande es la parte común de los mismos. El valor exacto de la dimensión de la suma se determina por el teorema siguiente:

TEOREMA. *La dimensión de la suma de dos subespacios lineales de un espacio* $\mathfrak{L}$ *es igual a la suma de las dimensiones de estos subespacios menos la dimensión de su intersección.*

Sean $\mathfrak{A}$ y $\mathfrak{B}$ los subespacios dados y sean r_1 y r_2 sus dimensiones respectivas. Sea m la dimensión de la intersección $\mathfrak{C}$ de estos subespacios. Tomemos en $\mathfrak{C}$ una base cualquiera $c_1, c_2, \ldots, c_m$. Los vectores $c_1, c_2, \ldots, c_m$ son linealmente independientes y pertenecen a $\mathfrak{A}$. Por esto, en $\mathfrak{A}$ se pueden encontrar unos vectores $a_1, a_2, \ldots, a_k$, tales, que el sistema $a_1, a_2, \ldots, a_k, c_1, \ldots, c_m$ sea una base de $\mathfrak{A}$ (véase el p. 4.2). Por esta misma razón, en el subespacio $\mathfrak{B}$ existen vectores $b_1, b_2, \ldots, b_p$ tales, que junto a los vectores $c_1, c_2, \ldots, c_m$ constituyen una base de $\mathfrak{B}$. Puesto que el número de vectores de una base coincide con la dimensión del espacio, entre los números k y p y las dimensiones de los subespacios $\mathfrak{A}$ y $\mathfrak{B}$ existen las relaciones

$$r_1 = k + m \quad \text{y} \quad r_2 = p + m.$$

Si demostramos que el sistema

$$a_1, \ldots, a_k, \quad c_1, \ldots, c_m, \quad b_1, \ldots, b_p \tag{2}$$

es una base del espacio $\mathfrak{A} + \mathfrak{B}$, el teorema 1 quedará demostrado, ya que la dimensión del espacio $\mathfrak{A} + \mathfrak{B}$ será igual a

$$k + m + p = r_1 + r_2 - m.$$

Todo vector a de $\mathfrak{A}$ se expresa linealmente en términos del sistema $a_1, \ldots, a_k, c_1, \ldots, c_m$, que constituye una base de $\mathfrak{A}$ y, por esto, se expresa también en términos del sistema (2). Análogamente, todo vector b de $\mathfrak{B}$ también se expresa linealmente en términos de (2). Pero en este caso la suma $a + b$, es decir, cualquier vector de $\mathfrak{A} + \mathfrak{B}$, se expresará linealmente en términos de (2). Resta demostrar que el sistema (2) es linealmente independiente.

Sea

$$\alpha_1 a_1 + \ldots + \alpha_k a_k + \gamma_1 c_1 + \ldots + \gamma_m c_m + \beta_1 b_1 + \ldots + \beta_p b_p = o \quad (3)$$

una relación de dependencia lineal entre estos vectores. Pongamos

$$b = \beta_1 b_1 + \ldots + \beta_p b_p.$$

El vector b se expresa linealmente en términos de los vectores $b_1, \ldots, b_p$ contenidos en el espacio $\mathfrak{B}$; luego, b pertenece a $\mathfrak{B}$. Por otro lado de (3) se desprende que

$$b = -\alpha_1 a_1 - \ldots - \alpha_k a_k - \gamma_1 c_1 - \ldots - \gamma_m c_m. \quad (4)$$

Puesto que $a_1, \ldots, a_k, c_1, \ldots, c_m$ están contenidos en $\mathfrak{A}$, de aquí se deduce que b también pertenece a $\mathfrak{A}$. Por consiguiente, b figura en la intersección de los espacios $\mathfrak{A}$ y $\mathfrak{B}$, es decir, la expresión (4) del vector b en términos de la base de $\mathfrak{A}$ no contiene términos con $a_1, \ldots, a_k$, en otras palabras,

$$\alpha_1 = \alpha_2 = \ldots = \alpha_k = 0.$$

Tomando estos valores en (3), obtenemos

$$\gamma_1 c_1 + \ldots + \gamma_m c_m + \beta_1 b_1 + \ldots + \beta_p b_p = o.$$

Pero el sistema $c_1, \ldots, c_m, b_1, \ldots, b_p$ es una base de $\mathfrak{B}$ y, por consiguiente, es linealmente independiente, de modo que

$$\gamma_1 = \ldots = \gamma_m = \beta_1 = \ldots = \beta_p = 0.$$

Hemos demostrado que el sistema (2) es linealmente independiente.

Del teorema demostrado se puede deducir una desigualdad que ofrece el valor mínimo de la dimensión de la intersección de unos subespacios. Consideremos unos subespacios lineales $\mathfrak{A}$ y $\mathfrak{B}$ de $\mathfrak{L}$ y sean r_1 y r_2 las dimensiones de estos subespacios, n la dimensión de $\mathfrak{L}$ y m la dimensión de la intersección $\mathfrak{A} \cap \mathfrak{B}$. En virtud del teorema, la dimensión de la suma $\mathfrak{A} + \mathfrak{B}$ es igual a $r_1 + r_2 - m$. Pero la dimensión de la suma $\mathfrak{A} + \mathfrak{B}$ es no mayor que la dimensión del espacio $\mathfrak{L}$. Por consiguiente, $r_1 + r_2 - m \leqslant n$, de donde se tiene $m \geqslant r_1 + r_2 - n$. Es decir, *la dimensión de la intersección de dos subespacios lineales del espacio $\mathfrak{L}$ no puede ser menor que el exceso de la suma de las dimensiones de estos subespacios respecto a la dimensión del espacio $\mathfrak{L}$.*

Por ejemplo, la intersección de dos planos del espacio de tres dimensiones contiene siempre una recta, la intersección de un subespacio de dos dimensiones con un subespacio de tres dimensiones en un espacio de cuatro dimensiones contiene una recta, la intersección de dos subespacios de tres dimensiones de un espacio de cuatro dimensiones contiene un plano, etc.

6.2. Sumas directas. Según la definición, todo vector a que pertenece a la suma de subespacios lineales

$$\mathfrak{A} = \mathfrak{A}_1 + \mathfrak{A}_2 + \ldots + \mathfrak{A}_s,$$

puede ser representado en la forma

$$a = a_1 + a_2 + \ldots + a_s \quad (a_i \in \mathfrak{A}_i, \ i = 1, \ \ldots, \ s). \tag{1}$$

Sin embargo, esta representación no será, en general, unívoca. Si todo vector a de $\mathfrak{A}$ admite sólo una representación de tipo (1), es decir, si de las igualdades (1) y de

$$a = a'_1 + a'_2 + \ldots + a'_s \quad (a'_i \in \mathfrak{A}_i, \ i = 1, \ \ldots, \ s) \tag{2}$$

se deduce que $a_1 = a'_1, \ \ldots, \ a_s = a'_s$, la suma se llama *suma directa* y se indica

$$\mathfrak{A} = \mathfrak{A}_1 \dot{+} \mathfrak{A}_2 \dot{+} \ldots \dot{+} \mathfrak{A}_s.$$

Las sumas directas poseen numerosas propiedades específicas, algunas de las cuales serán ahora consideradas. Ante todo observemos que en lugar de exigir que cualquier vector de la suma posee un desarrollo unívoco se puede exigir que el desarrollo unívoco exista sólo para el vector nulo: *si el vector nulo de la suma* $\mathfrak{A} = \mathfrak{A}_1 + \mathfrak{A}_2 + \ldots + \mathfrak{A}_s$ *admite un desarrollo único de tipo* (1), *es decir, si*

$$a_1 + a_2 + \ldots + a_s = o \quad (a_i \in \mathfrak{A}_i, \ i = 1, \ \ldots, \ s)$$

implica que $a_1 = a_2 = \ldots = a_s = o$, *la suma es directa.*

Para la demostración tomemos un vector cualquiera a de la suma y supongamos que admite dos desarrollos: (1) y (2). Restando un desarrollo del otro, obtenemos

$$(a_1 - a'_1) + (a_2 - a'_2) + \ldots + (a_s - a'_s) = o.$$

Puesto que aquí $(a_i - a'_i) \in \mathfrak{A}_i, \ i = 1, \ 2, \ \ldots, \ s$, tenemos según la hipótesis

$$a_1 - a'_1 = a_2 - a'_2 = \ldots = a_s - a'_s = o,$$

es decir,

$$a_1 = a'_1, \ a_2 = a'_2, \ \ldots, \ a_s = a'_s.$$

Hemos demostrado la proposición.

El teorema que sigue está relacionado con el empleo de los paréntesis en las sumas directas.

TEOREMA 1. *Si las descomposiciones de los subespacios lineales*

$$\mathfrak{A} = \mathfrak{A}_1 + \mathfrak{A}_2 + \ldots + \mathfrak{A}_s, \tag{3}$$

$$\left.\begin{aligned} \mathfrak{A}_1 &= \mathfrak{A}_{11} + \mathfrak{A}_{12} + \ldots + \mathfrak{A}_{1m_1}, \\ &\ldots\ldots\ldots\ldots \\ \mathfrak{A}_s &= \mathfrak{A}_{s1} + \mathfrak{A}_{s2} + \ldots + \mathfrak{A}_{sm_s}, \end{aligned}\right\} \tag{4}$$

son directas, la descomposición

$$\mathfrak{A} = \mathfrak{A}_{11} + \mathfrak{A}_{12} + \ldots + \mathfrak{A}_{1m_1} + \ldots + \mathfrak{A}_{s1} + \mathfrak{A}_{s2} + \ldots + \mathfrak{A}_{sm_s} \quad (5)$$

también es directa; en otras palabras, si en una suma directa todo sumando es sustituido por su descomposición directa, se obtendrá de nuevo una descomposición directa. Recíprocamente, si la descomposición (5) es directa, las descomposiciones (4) *y* (3) *también son directas.*

Demostremos primero la proposición directa. Sea

$$a_{11} + a_{12} + \ldots + a_{1m_1} + \ldots + a_{s1} + a_{s2} + \ldots + a_{sm_s} = o.$$

Escribamos esta igualdad en la forma

$$a_1 + a_2 + \ldots + a_s = o, \quad (6)$$

donde

$$a_i = a_{i1} + a_{i2} + \ldots + a_{im_i} \quad (i = 1, 2, \ldots, s). \quad (7)$$

Como que $a_i \in \mathfrak{A}_i$ y la suma (3) es directa, de (6) se deduce que $a_1 = a_2 = \ldots = a_s = o$ y la igualdad (7) se transforma en

$$a_{i1} + a_{i2} + \ldots + a_{im_i} = o \quad (i = 1, 2, \ldots, s). \quad (8)$$

Pero la descomposición

$$\mathfrak{A}_i = \mathfrak{A}_{i1} + \mathfrak{A}_{i2} + \ldots + \mathfrak{A}_{im_i}$$

es, por hipótesis, directa y por ello de (8) se deduce que $a_{i1} = a_{i2} = \ldots = a_{im_i} = o$. Por consiguiente, la suma (5) es directa que es lo que se quería demostrar. Repitiendo los razonamientos en el orden contrario, obtendremos la demostración de la afirmación recíproca.

El teorema 1 permite considerar la suma directa de varios subespacios como el resultado de sucesivas sumas directas de dos sumandos. Las condiciones que garantizan que la suma de dos subespacios sea directa pueden ser enunciadas de la siguiente forma conveniente.

TEOREMA 2. *Para que la suma de dos subespacios lineales de un espacio* $\mathfrak{L}$ *sea directa es necesario y suficiente que la intersección de estos subespacios sea nula.*

En efecto, si la intersección de dos subespacios $\mathfrak{A}$ y $\mathfrak{B}$ contiene un vector no nulo a, para el vector o se puede escribir la descomposición

$$a + (-a) = o,$$

donde $a \neq o$, $a \in \mathfrak{A}$ y $-a \in \mathfrak{B}$, y, por consiguiente, la suma $\mathfrak{A} + \mathfrak{B}$ no será directa. Viceversa, si la suma $\mathfrak{A} + \mathfrak{B}$ no es directa, para el vector nulo deberá existir la descomposición

$$a + b = o \quad (a \neq o,\ a \in \mathfrak{A} \text{ y } b \in \mathfrak{B}).$$

Puesto que $b = -a$, el vector b figura también en el subespacio $\mathfrak{A}$. Por consiguiente, la intersección $\mathfrak{A} \cap \mathfrak{B}$ contiene, en este caso, un vector no nulo b, que es lo que se quería demostrar.

TEOREMA 3. *La dimensión de una suma directa de subespacios es igual a la suma de las dimensiones de estos subespacios.*

Si tenemos dos sumandos, entonces la dimensión de la suma es igual, de acuerdo con el p. 6.1, a la suma de las dimensiones de los sumandos menos la dimensión de la intersección. Pero, según el teorema 2, la intersección de subespacios en el caso de una suma directa es nula y su dimensión es igual a cero. Por esto, *la dimensión de una suma directa de dos subespacios es igual a la suma de sus dimensiones.* Si se tiene más de dos sumandos, la demostración se realiza fácilmente por inducción.

Del teorema 3 se desprende el siguiente corolario: *si el subespacio $\mathfrak{A}$ es la suma directa de los subespacios $\mathfrak{A}_1, \ldots, \mathfrak{A}_s$, entonces, escogiendo en cada subespacio $\mathfrak{A}_i$ una base $a_{i1}, a_{i2}, \ldots, a_{im_i}$ $(i = 1, 2, \ldots, s)$ y uniendo estas bases en un sistema*

$$a_{11}, a_{12}, \ldots, a_{1m_1}, \ldots, a_{s1}, a_{s2}, \ldots, a_{sm_s}, \tag{9}$$

obtendremos una base del subespacio $\mathfrak{A}$.

Efectivamente, todo vector del subespacio $\mathfrak{A}_i$ se expresa linealmente en términos de los vectores (9) y por esto cualquier vector de $\mathfrak{A}$ también se expresará linealmente en términos de los vectores (9). En virtud del teorema 3, el número de vectores del sistema (9) es igual a la dimensión del subespacio $\mathfrak{A}$. Luego, el sistema (9) es una base de $\mathfrak{A}$, que es lo que se quería demostrar.

El teorema 3 admite la siguiente inversión: *si la dimensión de una suma de subespacios lineales es igual a la suma de las dimensiones de los mismos, la suma es directa.*

Sea primero $\mathfrak{A} = \mathfrak{A}_1 + \mathfrak{A}_2$ y sea

$$\dim. \mathfrak{A} = \dim. \mathfrak{A}_1 + \dim. \mathfrak{A}_2.$$

En virtud del teorema del p. 6.1, de aquí se deduce que la dimensión de la intersección $\mathfrak{A}_1 \cap \mathfrak{A}_2$ es igual a cero, es decir, que esta intersección es nula. A su vez, de aquí se deduce, de acuerdo con el teorema 2, que la suma $\mathfrak{A}_1 + \mathfrak{A}_2$ es directa. Si el número de sumandos es mayor que dos, es suficiente aplicar para la demostración la inducción.

6.3. Sistemas de ecuaciones lineales homogéneas. Con el fin de ilustrar la teoría de subespacios lineales consideremos el problema de resolución de los sistemas de ecuaciones lineales homogéneas, es decir, de los sistemas de tipo

$$\left.\begin{aligned} \xi_1\alpha_{11} + \xi_2\alpha_{21} + \ldots + \xi_n\alpha_{n1} &= 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots & \\ \xi_1\alpha_{1m} + \xi_2\alpha_{2m} + \ldots + \xi_n\alpha_{nm} &= 0. \end{aligned}\right\} \tag{1}$$

donde α_{ij} son elementos de un cuerpo K, mientras que $\xi_1, \ldots, \xi_n$ son las incógnitas; los valores de las últimas se buscan en el cuerpo K. El sistema (1) puede ser representado en la siguiente forma matricial

$$x \cdot A = O, \tag{2}$$

donde $x = [\xi_1 \ldots, \xi_n]$ es la fila de incógnitas y $A = \|\alpha_{ij}\|$ es la matriz en la que la j-ésima columna está formada por los coeficientes de la j-ésima ecuación del sistema (1). La fila nula $[0, \ldots, 0]$ ofrece una solución del sistema (1) cualesquiera que sean los coeficientes α_{ij}. Esta fila se llama solución *nula* o *trivial* del sistema (1). La existencia de la solución trivial indica que un sistema de ecuaciones lineales homogéneas nunca es contradictorio.

Para poder estudiar con más detalle las propiedades de las soluciones del sistema (1) indiquemos por $\mathfrak{L}$ el espacio de filas de longitud n sobre el cuerpo K y consideremos toda solución $[\xi_1^0, \ldots, \xi_n^0]$ del sistema (1) como un vector x del espacio $\mathfrak{L}$ que satisface la ecuación (2). *El conjunto de todas las soluciones de la ecuación* (2) *es un subespacio lineal del espacio* $\mathfrak{L}$. Efectivamente, de (2) y de $y \cdot A = O$ se deduce que

$$(\lambda x + \mu y)\, A = \lambda\,(xA) + \mu\,(yA) = O$$

cualesquiera que sean $\lambda, \mu \in K$. En otras palabras, una combinación lineal arbitraria de soluciones del sistema (1) es también una solución del sistema (1). Puesto que el espacio $\mathfrak{L}$ es de dimensión n, el espacio de soluciones del sistema (1) es de dimensión no mayor que n.

Se dice que unas soluciones $x_1, \ldots, x_s$ del sistema (2) forman *un sistema fundamental de soluciones* de (2), si cualquier solución del sistema (2) puede ser representada en forma de una combinación lineal $\lambda_1 x_1 + \ldots + \lambda_s x_s$ de las soluciones indicadas y, al mismo tiempo, ninguna de las soluciones $x_1, \ldots, x_s$ puede ser representada como una combinación lineal de las restantes. En términos de la teoría de espacios lineales esto significa que el conjunto $x_1, \ldots, x_s$ es simplemente *una base del subespacio de soluciones* y que el número de soluciones fundamentales es la dimensión del subespacio de soluciones.

El rango de la matriz A se llama *rango del sistema homogéneo* (1). Los sistemas homogéneos de tipo (1) representan un caso particular de los sistemas lineales generales estudiados ya en el p. 5.3. Empleando el teorema de Kronecker—Capelli demostrado en aquella ocasión se obtiene fácilmente el siguiente teorema principal:

TEOREMA SOBRE LAS ECUACIONES LINEALES HOMOGÉNEAS. *La dimensión del espacio de soluciones del sistema* (1) *de ecuaciones lineales homogéneas con n incógnitas es igual a la diferencia* $n - r$, *donde r es el rango del sistema* (1).

Si $r = n$, el sistema (1) debe tener, según el teorema de Kronecker—Capelli, una solución única que es, por consiguiente, la solución nula. El subespacio de soluciones consta solamente del vector nulo y su dimensión es igual a 0, es decir, coincide con la diferencia $n-r$.

Sea $r < n$. En virtud del teorema de Kronecker—Capelli, el sistema (1) será en este caso (después de realizar de nuevo una enumeración adecuada de las incógnitas y de las ecuaciones) equivalente a un sistema de tipo

$$\left.\begin{aligned} \xi_1 &= \xi_{r+1}\gamma_{r+1,\,1} + \ldots + \xi_n\gamma_{n1}, \\ &\ldots\ldots\ldots\ldots \\ \xi_r &= \xi_{r+1}\gamma_{r+1,\,r} + \ldots + \xi_n\gamma_{nr}, \end{aligned}\right\} \tag{3}$$

donde γ_{sl} son determinados elementos de K. A las incógnitas $\xi_{r+1}, \ldots, \xi_n$ se les puede asignar aquí cualesquiera valores de K. Tomando sucesivamente

$$\begin{aligned} &[\xi_{r+1}, \xi_{r+2}, \ldots, \xi_n] = [1, 0, \ldots, 0], \\ &[\xi_{r+1}, \xi_{r+2}, \ldots, \xi_n] = [0, 1, \ldots, 0], \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ &[\xi_{r+1}, \xi_{r+2}, \ldots, \xi_n] = [0, 0, \ldots, 1] \end{aligned}$$

y calculando toda vez los valores de las incógnitas $\xi_1, \ldots, \xi_n$ según las fórmulas (3), encontramos el siguiente sistema de soluciones de (1):

$$\begin{aligned} c_1 &= [\gamma_{r+1,\,1}, \ldots, \gamma_{r+1,\,r}, 1, 0, \ldots, 0], \\ c_2 &= [\gamma_{r+2,\,1}, \ldots, \gamma_{r+2,\,r}, 0, 1, \ldots, 0], \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ c_{n-r} &= [\gamma_{n1}, \ldots, \gamma_{nr}, \quad 0, 0, \ldots, 1]. \end{aligned} \tag{4}$$

Sabemos que el número de filas linealmente independientes en el sistema $c_1, \ldots, c_{n-r}$ coincide con el rango de la matriz formada por estas filas. Pero de (4) se ve que la matriz indicada contiene un menor, formado por las últimas $n-r$ columnas, que es la matriz unidad. Por esto, el rango de la matriz es igual a $n-r$ y, por consiguiente, todos los vectores $c_1, \ldots, c_{n-r}$ son linealmente independientes. Por otro lado, las fórmulas (3) muestran que toda solución $x = [\xi_1^0, \ldots, \xi_n^0]$ satisface la relación

$$x = \xi_{r+1}^0 c_1 + \ldots + \xi_n^0 c_{n-r},$$

es decir, es una combinación lineal de las soluciones $c_1, \ldots, c_{n-r}$. Hemos verificado de esta forma que el sistema $c_1, \ldots, c_{n-r}$ es una base del espacio de soluciones del sistema (1) y, por esto, la dimensión de dicho espacio es igual a $n-r$.

Indiquemos dos corolarios directos del teorema demostrado.

COROLARIO 1. *Para que un sistema de n ecuaciones lineales homogéneas de n incógnitas*

$$\left.\begin{aligned} \xi_1\alpha_{11} + \ldots + \xi_n\alpha_{n1} &= 0, \\ \ldots\ldots\ldots\ldots & \\ \xi_1\alpha_{n1} + \ldots + \xi_n\alpha_{nn} &= 0 \end{aligned}\right\}$$

con coeficientes de un cuerpo cualquiera K no tenga soluciones no nulas, es necesario y suficiente que la matriz de esta sistema

$$A = \begin{vmatrix} \alpha_{11} & \ldots & \alpha_{n1} \\ \ldots & \ldots & \ldots \\ \alpha_{1n} & \ldots & \alpha_{nn} \end{vmatrix}$$

sea invertible.

Efectivamente, la condición $n = r$ significa que el rango de la matriz A debe coincidir con su orden, es decir, la matriz A debe ser invertible.

COROLARIO 2. *Para que un sistema de n ecuaciones lineales homogéneas de n incógnitas con coeficientes de un cuerpo conmutativo tenga solución no nula es necesario y suficiente que el determinante de la matriz de este sistema sea igual a cero.*

En efecto, una matriz cuadrada formada por elementos de un cuerpo conmutativo no es invertible si, y sólo si, su determinante es igual a cero.

Hemos visto que el teorema sobre las ecuaciones lineales homogéneas es un corolario directo del teorema de Kronecker—Capelli. El último teorema, además de ser cierto para los sistemas lineales con coeficientes a la derecha, tiene lugar también para los sistemas lineales con coeficientes a la izquierda. Por esto, junto al teorema sobre las ecuaciones lineales homogéneas y al corolario 1, son válidas las proposiciones análogas relacionadas con sistemas de ecuaciones lineales homogéneas con coeficientes a la izquierda, a saber: el conjunto de soluciones columnas $x = [\xi_1^0, \ldots, \xi_n^0]$ de un sistema de ecuaciones lineales homogéneas

$$\left.\begin{aligned} \alpha_{11}\xi_1 + \ldots + \alpha_{1n}\xi_n &= 0, \\ \ldots\ldots\ldots\ldots & \\ \alpha_{m1}\xi_1 + \ldots + \alpha_{mn}\xi_n &= 0 \end{aligned}\right\} \qquad (5)$$

con coeficientes de un cuerpo K es un subespacio lineal del espacio formado por todas las columnas de longitud n sobre K. La dimensión de este espacio de soluciones es igual a $n-r$, donde r es el rango de la matriz $A = \|\alpha_{ij}\|$, en la que la i-ésima fila está formada por los coeficientes de la i-ésima ecuación del sistema (5).

Ejemplos y problemas

1. Hállese la dimensión del subespacio lineal tendido sobre los vectores $a=[1, 3, 2, 1]$, $b=[4, 9, 5, 4]$ y $c=[3, 7, 4, 3]$.

2. Si $e_1, e_2, \ldots, e_n$ es una base de un espacio lineal $\mathfrak{L}$ y $\mathfrak{A}_i$ es el subespacio lineal tendido sobre e_i, se tiene

$$\mathfrak{L}=\mathfrak{A}_1 \dot{+} \mathfrak{A}_2 \dot{+} \ldots \dot{+} \mathfrak{A}_n.$$

3. Para todo subespacio propio $\mathfrak{A}$ de un espacio $\mathfrak{L}$ existe un subespacio lineal $\mathfrak{B}$, tal, que $\mathfrak{L}=\mathfrak{A} \dot{+} \mathfrak{B}$.

4. Demuéstrese la siguiente generalización del teorema 2 del p. 6.2. para que la suma de varios subespacios lineales dados sea directa es necesario y suficiente que cada uno de los subespacios dados tenga intersección nula con la suma de los restantes.

5. Demuéstrese que la intersección de todos los subespacios lineales de un espacio $\mathfrak{L}$, que contienen a los subespacios lineales dados, es igual a la suma de los últimos.

6. Demuéstrese que para cualesquiera subespacios $\mathfrak{A}$, $\mathfrak{B}$ y $\mathfrak{C}$ de un espacio lineal tienen lugar las igualdades

$$\mathfrak{A}(\mathfrak{A}\mathfrak{B}+\mathfrak{C})=\mathfrak{A}\mathfrak{B}+\mathfrak{A}\mathfrak{C},$$
$$(\mathfrak{A}+\mathfrak{B})(\mathfrak{A}+\mathfrak{C})=\mathfrak{A}+(\mathfrak{A}+\mathfrak{B})\mathfrak{C},$$

donde, para abreviar, la intersección de subespacios se indica igual que el producto.

7. Todo espacio lineal $\mathfrak{L}$ de dimensión infinita contiene un subespacio lineal propio cuya dimensión coincide con la dimensión de todo el espacio $\mathfrak{L}$.

8. ¿Para qué valores de λ (del cuerpo de los números complejos) el sistema de ecuaciones

$$\left.\begin{array}{r} x+2\lambda y-\lambda z=y, \\ 2x+\ y-\ z=\lambda x, \\ x+\ \lambda y-\ z=0 \end{array}\right\}$$

tiene solución no nula? ¿Para qué valores de λ el espacio de soluciones de este sistema es de mayor dimensión?

9. ¿Para qué valores del parámetro λ el sistema de ecuaciones

$$\left.\begin{array}{r} x+\ y+(\lambda-2)z=1, \\ \lambda x+3y+\qquad \lambda z=2 \end{array}\right\}$$

es compatible?

10. ¿Qué dimensión tienen los espacios de soluciones de los sistemas

$$\left.\begin{array}{r} x-iy+\quad iz=0 \\ jx+ky-\quad z=0 \\ x-iy+(2i-j)z=0 \end{array}\right\} \quad \text{y} \quad \left.\begin{array}{r} x-yi+zi\quad =0 \\ xj+yk-z\quad =0 \\ x-yi+z(2i-j)=0 \end{array}\right\}$$

considerados sobre el cuerpo de los cuaternios (p.1.5)?

11. ¿Existe un polinomio en variables cuaternias a, b, c y d y con coeficientes cuaternios, tal, que su anulación sea la condición necesaria y suficiente para que el sistema de ecuaciones lineales

$$ax+by=0 \quad \text{y} \quad cx+dy=0$$

tenga solución no nula?

12. ¿Existe un polinomio en $a, \bar{a}, b, \bar{b}, c, \bar{c}, d$ y $\bar{d}$ que satisfaga las exigencias del ejercicio anterior?

Capítulo III Aplicaciones lineales

El objetivo principal de este capítulo es el estudio de las propiedades de las aplicaciones de espacios lineales. En los parágrafos 8 y 10, así como en los puntos 9.1, 11.1 y 11.2, el cuerpo principal no se somete a ninguna restricción. En los puntos 9.2, 9.3, 11.3 y 11.4 y también en el § 12 se supone que el cuerpo principal es un cuerpo conmutativo. En el § 7, de carácter de introducción, se establece una serie de conceptos y de propiedades relacionados con aplicaciones de conjuntos absolutamente arbitrarios.

§ 7. Aplicaciones de conjuntos arbitrarios

7.1. Producto de aplicaciones. Consideremos un conjunto $\mathfrak{M}$ de entes arbitrarios. Este conjunto puede estar compuesto tanto por un número finito como infinito de elementos. Se llama *aplicación* del conjunto $\mathfrak{M}$ toda ley que permite a partir de cualquier elemento del conjunto $\mathfrak{M}$ encontrar de nuevo un elemento de $\mathfrak{M}$. Convendremos en indicar las aplicaciones por las letras $\mathcal{A}$, $\mathcal{B}$, ... Si m es un elemento de $\mathfrak{M}$, indicaremos por $m\mathcal{A}$ aquel elemento del conjunto $\mathfrak{M}$ que se obtiene del elemento m mediante la aplicación $\mathcal{A}$. El elemento $m\mathcal{A}$ se denomina *imagen* del elemento m en la aplicación $\mathcal{A}$, mientras que m se denomina *imagen recíproca* del elemento $m\mathcal{A}$.

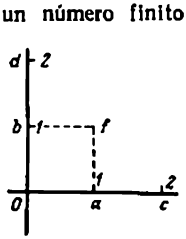


Fig. 2.

Consideremos, a título de ejemplo, el conjunto de todos los puntos del plano y sea $\mathcal{U}$ la aplicación que consiste en el giro de los puntos de este plano alrededor de uno de sus puntos O en 90° en contra del movimiento de las agujas del reloj (fig. 2).

Recurriendo a la figura, vemos que

$$a\mathcal{U} = b \quad \text{y} \quad c\mathcal{U} = d.$$

Análogamente, si $\mathcal{D}$ significa el traslado de los puntos en una unidad paralelamente al eje Oa, se tiene.

$$a\mathcal{D} = c \quad \text{y} \quad b\mathcal{D} = f.$$

Dos aplicaciones $\mathcal{A}$ y $\mathcal{B}$ del conjunto $\mathfrak{M}$ se llaman *iguales*, si para todo elemento m de $\mathfrak{M}$ se tiene.

$$m\mathcal{A} = m\mathcal{B}.$$

Uno de los conceptos principales de la teoría de aplicaciones es el de producto de aplicaciones que se introduce del modo siguiente. Sean $\mathcal{A}$ y $\mathcal{B}$ dos aplicaciones del conjunto $\mathfrak{M}$. La primera transforma un elemento arbitrario m del conjunto $\mathfrak{M}$ en $m\mathcal{A}$. Si este nuevo elemento se somete a la aplicación $\mathcal{B}$, se obtendrá el elemento $(m\mathcal{A})\mathcal{B}$. La aplicación que transforma el elemento m directamente en $(m\mathcal{A})\mathcal{B}$ se llama *producto* de $\mathcal{A}$ por $\mathcal{B}$ y se designa $\mathcal{A}\mathcal{B}$. Es decir, se toma por definición que

$$m(\mathcal{A}\mathcal{B}) = (m\mathcal{A})\mathcal{B}.$$

Si al elemento m se aplica primero la aplicación $\mathcal{B}$ y después la aplicación $\mathcal{A}$, se obtendrá el elemento $(m\mathcal{B})\mathcal{A}$ que puede no coincidir con $(m\mathcal{A})\mathcal{B}$. Efectivamente, en el ejemplo con el plano, considerado anteriormente, se tiene

$$a(\mathcal{U}\mathcal{D}) = (a\mathcal{U})\mathcal{D} = b\mathcal{D} = f,$$
$$a(\mathcal{D}\mathcal{U}) = (a\mathcal{D})\mathcal{U} = c\mathcal{U} = d$$

y, por consiguiente, $\mathcal{U}\mathcal{D} \neq \mathcal{D}\mathcal{U}$. Por lo tanto, *el producto de aplicaciones depende, en general, del orden de los factores*. Vemos, pues, que una de las leyes principales que se verifica para el producto de los números no se cumple para las aplicaciones. Sin embargo, la otra ley principal—*la asociatividad de la multiplicación*—se conserva para las aplicaciones. En efecto, sean $\mathcal{A}$, $\mathcal{B}$ y $\mathcal{C}$ aplicaciones arbitrarias del conjunto $\mathfrak{M}$ y sea m uno de sus elementos. Por definición, tenemos

$$m((\mathcal{A}\mathcal{B})\mathcal{C}) = (m(\mathcal{A}\mathcal{B}))\mathcal{C} = ((m\mathcal{A})\mathcal{B})\mathcal{C},$$
$$m(\mathcal{A}(\mathcal{B}\mathcal{C})) = (m\mathcal{A})(\mathcal{B}\mathcal{C}) = ((m\mathcal{A})\mathcal{B})\mathcal{C},$$

de donde

$$(\mathcal{A}\mathcal{B})\mathcal{C} = \mathcal{A}(\mathcal{B}\mathcal{C}).$$

Empleando esta ley es fácil deducir que *el producto de cualquier número finito de aplicaciones, tomadas en un orden determinado, no depende de la disposición de los paréntesis*. Así, por ejemplo,

$$(\mathcal{A}\mathcal{B})(\mathcal{C}\mathcal{D}) = ((\mathcal{A}\mathcal{B})\mathcal{C})\mathcal{D} = (\mathcal{A}(\mathcal{B}\mathcal{C}))\mathcal{D} = \mathcal{A}((\mathcal{B}\mathcal{C})\mathcal{D}).$$

Por esto, en los productos que contienen varios factores se pueden omitir los paréntesis y se puede hablar simplemente del producto de dos, de tres y de un número mayor de aplicaciones. El producto

de n factores iguales a $\mathcal{A}$ se llama n-ésima potencia de la aplicación $\mathcal{A}$ y se indica por $\mathcal{A}^n$. Las operaciones con las potencias se realizan siguiendo las reglas corrientes

$$\mathcal{A}^m\mathcal{A}^n = \mathcal{A}^{m+n}, \tag{1}$$

$$(\mathcal{A}^m)^n = \mathcal{A}^{mn}; \tag{2}$$

la demostración de las mismas es evidente.

Si el producto de dos aplicaciones $\mathcal{A}$ y $\mathcal{B}$ no depende del orden de los factores, se dice que $\mathcal{A}$ y $\mathcal{B}$ son *permutables* o que *conmutan*. La fórmula (1) señala que *las potencias de una misma aplicación son permutables*.

Si las aplicaciones $\mathcal{A}$ y $\mathcal{B}$ conmutan, se tiene

$$(\mathcal{AB})^2 = \mathcal{ABAB} = \mathcal{AABB} = \mathcal{A}^2\mathcal{B}^2$$

y, en general,

$$(\mathcal{AB})^n = \mathcal{A}^n\mathcal{B}^n. \tag{3}$$

En cambio, si $\mathcal{A}$ y $\mathcal{B}$ no conmutan, la fórmula (3) puede no tener lugar.

7.2. Las aplicaciones idéntica e inversa. Entre todas las aplicaciones de un conjunto $\mathfrak{M}$ desempeña un papel especial la aplicación que pone en correspondencia a todo elemento m de este conjunto el propio elemento m. Esta aplicación lleva el nombre de aplicación *unidad* o *idéntica* y será indicada en adelante por $\mathcal{E}$. Es decir, para todo m se tiene

$$m\mathcal{E} = m.$$

Sea $\mathcal{A}$ una aplicación arbitraria del conjunto $\mathfrak{M}$. Puesto que

$$m(\mathcal{E}\mathcal{A}) = (m\mathcal{E})\mathcal{A} = m\mathcal{A}$$

y

$$m(\mathcal{A}\mathcal{E}) = (m\mathcal{A})\mathcal{E} = m\mathcal{A},$$

se tiene

$$\mathcal{E}\mathcal{A} = \mathcal{A}\mathcal{E} = \mathcal{A}.$$

Si dada la aplicación $\mathcal{A}$ se puede hallar una aplicación $\mathcal{B}$ tal que

$$\mathcal{AB} = \mathcal{BA} = \mathcal{E}, \tag{4}$$

se dice que $\mathcal{B}$ es la *inversa* de $\mathcal{A}$, mientras que $\mathcal{A}$ se denomina *invertible*. Es fácil ver que *toda aplicación invertible tiene sólo una inversa*. Efectivamente, si $\mathcal{A}$ posee dos aplicaciones inversas $\mathcal{B}$ y $\mathcal{C}$, entonces, multiplicando la relación

$$\mathcal{AC} = \mathcal{E}$$

a la izquierda por $\mathcal{B}$ y empleando las igualdades (4) y la asociatividad de la multiplicación de las aplicaciones, obtenemos $\mathcal{BC} = \mathcal{BB}$, es decir, $\mathcal{C} = \mathcal{B}$.

La aplicación inversa de $\mathcal{A}$ se indica por $\mathcal{A}^{-1}$. Las relaciones (4) son simétricas respecto a $\mathcal{A}$ y $\mathcal{B}$; por esto, si $\mathcal{B}$ es la inversa de $\mathcal{A}$ resulta que $\mathcal{A}$ es la inversa de $\mathcal{B}$, es decir,

$$(\mathcal{A}^{-1})^{-1} = \mathcal{A}. \tag{5}$$

Tomemos por definición

$$\mathcal{A}^0 = \mathcal{E} \quad \text{y} \quad \mathcal{A}^{-n} = (\mathcal{A}^{-1})^n \quad (n = 1, 2, 3, \ldots).$$

De (4) y (5) se deduce fácilmente que las fórmulas (1) y (2) son válidas no sólo para exponentes enteros positivos, sino para todos los exponentes enteros. En particular,

$$(\mathcal{A}^n)^{-1} = (\mathcal{A}^{-1})^n = \mathcal{A}^{-n}.$$

Además, la relación

$$\mathcal{AB} \cdot \mathcal{B}^{-1}\mathcal{A}^{-1} = \mathcal{E}$$

implica que

$$(\mathcal{AB})^{-1} = \mathcal{B}^{-1}\mathcal{A}^{-1}.$$

7.3. Aplicaciones biyectivas. No toda aplicación es invertible. El teorema que sigue ofrece un criterio simple de aplicaciones invertibles.

TEOREMA *Para que una aplicación $\mathcal{A}$ de un conjunto $\mathfrak{M}$ sea invertible es necesario y suficiente que $\mathcal{A}$ sea una aplicación biyectiva del conjunto $\mathfrak{M}$ sobre sí mismo, es decir, que para todo elemento de $\mathfrak{M}$ exista en $\mathfrak{M}$ su imagen recíproca y que distintos elementos de $\mathfrak{M}$ se transformen por la aplicación $\mathcal{A}$ en distintos elementos.*

Demostremos primero la necesidad. Supongamos que $\mathcal{A}$ posee la aplicación inversa $\mathcal{B}$, de manera que

$$\mathcal{AB} = \mathcal{BA} = \mathcal{E}.$$

Tomemos en $\mathfrak{M}$ un elemento cualquiera m y sea $m\mathcal{B} = n$. Multiplicando esta relación por $\mathcal{A}$ y sustituyendo $\mathcal{BA}$ por $\mathcal{E}$, obtendremos $m = n\mathcal{A}$, es decir, todo elemento m de $\mathfrak{M}$ es la imagen de un elemento n de $\mathfrak{M}$. Por otro lado, si dos elementos m_1 y m_2 son transformados por la aplicación $\mathcal{A}$ en un mismo elemento

$$m_1\mathcal{A} = m_2\mathcal{A},$$

entonces, multiplicando esta relación por $\mathcal{B}$, obtenemos $m_1 = m_2$. Por consiguiente, todo elemento de $\mathfrak{M}$ tiene en $\mathfrak{M}$ sólo una imagen recíproca.

Demostremos ahora la suficiencia de las condiciones. Según éstas, para todo elemento m del conjunto $\mathfrak{M}$ existe un elemento, y sólo uno, n tal que

$$n\mathcal{A} = m \tag{6}$$

Indiquemos por $\mathcal{B}$ la aplicación que transforma m en n. Es decir,

$$m\mathcal{B} = n. \tag{7}$$

Multiplicando (7) por $\mathcal{A}$ y empleando la igualdad (6), obtenemos

$$m(\mathcal{B}\mathcal{A}) = n\mathcal{A} = m.$$

Puesto que m es un elemento arbitrario, de aquí se tiene $\mathcal{B}\mathcal{A} = \mathcal{E}$. Análogamente, multiplicando (6) por $\mathcal{B}$ y empleando (7), obtenemos $\mathcal{A}\mathcal{B} = \mathcal{E}$. Por consiguiente, $\mathcal{B}$ es la aplicación inversa deseada.

7.4. Sustituciones. Las aplicaciones de conjuntos finitos se representan, generalmente, mediante tablas, colocando en la primera fila de las mismas los símbolos de los elementos del conjunto dado en cierto orden y debajo de ellos los símbolos de los elementos que les *corresponden*. Por ejemplo,

$$\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$$

es aquella aplicación del conjunto de los números 1, 2, y 3 en la que el 1 se transforma en el 3, el 2 en el 1 y el 3 en el 2 o, empleando la notación aceptada anteriormente,

$$1\sigma = 3, \quad 2\sigma = 1 \quad \text{y} \quad 3\sigma = 2.$$

Las aplicaciones invertibles de conjuntos finitos se llaman *sustituciones*. Para que una aplicación representada en forma de una tabla sea una sustitución es necesario y suficiente que tanto en la fila superior como en la inferior aparezcan los símbolos de todos los elementos del conjunto, con la particularidad de que todo elemento figure sólo una vez. Para representar la sustitución inversa σ^{-1} basta, evidentemente, convertir la fila inferior de la tabla de σ en la superior y la superior en la inferior.

Sea $F(x_1, \ldots, x_n)$ una función de las variables $x_1, \ldots, x_n$ y sea σ una sustitución de los números $1, \ldots, n$. El resultado de la sustitución σ aplicada a las variables $x_1, \ldots, x_n$ de la función F es, por definición, la expresión

$$F\sigma = F(x_{1\sigma}, \ldots, x_{n\sigma}).$$

De esta definición se deduce directamente que para cualesquiera sustituciones ρ y σ y para cualesquiera funciones F y G de $x_1, \ldots, x_n$ se tiene

$$F(\rho\sigma) = (F\rho)\sigma, \tag{8}$$

$$(FG)\sigma = F\sigma \cdot G\sigma \quad \text{y} \quad (F+G)\sigma = F\sigma + G\sigma. \tag{9}$$

Tomemos ahora para la función F la expresión

$$\Delta = \prod_{i<j} (x_i - x_j) = (x_1 - x_2)(x_1 - x_3)\ldots(x_1 - x_n)(x_2 - x_3)\ldots$$
$$\ldots(x_{n-1} - x_n). \tag{10}$$

Está claro que para cualquier sustitución σ se tiene $\Delta\sigma = \pm\Delta$. Si $\Delta\sigma = \Delta$, se dice que σ es una sustitución *par* y, si $\Delta\sigma = -\Delta$, se dice que σ es *impar*. Para toda aplicación no invertible σ, tenemos $\Delta\sigma = 0$. Por esto, para cualquier aplicación σ del conjunto de los números 1, ..., n tenemos

$$\Delta\sigma = \varepsilon_\sigma \Delta,$$

donde $\varepsilon_\sigma = +1$, si σ es una sustitución par, $\varepsilon_\sigma = -1$, si σ es una sustitución impar, y $\varepsilon_\sigma = 0$, si σ es una aplicación no invertible. El valor del símbolo ε_σ se denomina también *signatura* de la aplicación σ.

Para dos aplicaciones arbitrarias ρ y σ tenemos de las fórmulas (8) y (9)

$$\varepsilon_{\rho\sigma}\Delta = \Delta(\rho\sigma) = (\Delta\rho)\sigma = \varepsilon_\rho \varepsilon_\sigma \Delta,$$

de donde

$$\varepsilon_{\rho\sigma} = \varepsilon_\rho \varepsilon_\sigma,$$

es decir, *la signatura del producto de aplicaciones es igual al producto de las signaturas de los factores*. Puesto que la signatura de la aplicación idéntica es, evidentemente, igual a +1, resulta que *la signatura de la sustitución inversa siempre coincide con la signatura de la sustitución dada*.

Se llama *sustitución cíclica*, o *ciclo*, $(i_1 i_2 \ldots i_m)$ la sustitución σ en la que $i_1\sigma = i_2$, $i_2\sigma = i_3$, ..., $i_{m-1}\sigma = i_m$, $i_m\sigma = i_1$ e $i\sigma = i$ para los demás elementos i del conjunto, si es que existen. En particular, se llama *ciclo doble*, o *trasposición*, (ij) la sustitución que cambie los elementos i y j y no altere los demás elementos. Cambiando los índices 1 y 2 en la expresión (10) para Δ, veremos fácilmente que $\Delta(1\ 2) = -\Delta$, es decir, que el ciclo doble (1 2) es una sustitución impar. Por otro lado, mediante cálculo directo se comprueba que

$$(1\ i)(1\ 2)(1\ i) = (i\ 2) \quad \text{y} \quad (2\ j)(i\ 2)(2\ j) = (i\ j)$$

y puesto que, además, para cualesquiera ρ y σ siempre $\varepsilon_{\rho\sigma\rho} = \varepsilon_\sigma$, obtenemos

$$\varepsilon_{(ij)} = \varepsilon_{(2j)(i2)(2j)} = \varepsilon_{(i2)} = \varepsilon_{(12)},$$

es decir, *todo ciclo doble es una sustitución impar*.

De aquí se deduce, además, que el producto de un número impar de ciclos dobles es una sustitución impar y que el producto de un número par de ciclos dobles es una sustitución par. En particular, de la fórmula

$$(1\ 2\ 3 \ldots m) = (1\ 2)(1\ 3) \ldots (1\ m),$$

que se comprueba fácilmente, se deduce que un ciclo $(i_1\ i_2 \ldots i_m)$ de

longitud par m es una sustitución impar y que un ciclo de longitud impar es una sustitución par.

Es claro que los ciclos sin elementos comunes representan sustituciones que conmutan. Al mismo tiempo, cualquier sustitución puede ser descompuesta fácilmente en un producto de ciclos sin elementos comunes. Para ello se toma un elemento cualquiera i_1 del conjunto y se mira en que elemento i_s lo transforma la sustitución. Si resulta que $i_1 = i_s$, obtenemos para el primer factor el ciclo unidad (i_1), es decir, la sustitución idéntica. Si resulta que $i_1 \neq i_s$, tomamos la imágen i_2 del elemento i_s. Para $i_2 = i_1$ obtenemos el factor en forma del ciclo $(i_1\ i_2)$; si $i_2 \neq i_1$, pasamos a considerar la imagen i_4 del elemento i_2. Si resulta que $i_4 = i_1$, el ciclo se cierra y se obtiene el factor $(i_1\ i_2\ i_3)$. etc.; considerando uno tras otro todos los elementos del conjunto, obtenemos la descomposición de la sustitución en ciclos sin elementos comunes. Por ejemplo, tenemos

$$\rho = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 2 & 1 \end{pmatrix} = (1\ 3\ 5)(2\ 4).$$

Multiplicando las signaturas de los ciclos obtenemos que $\varepsilon_\rho = -1$.

Existe también otra forma para determinar la signatura de una sustitución: mediante el cálculo del número de las así llamadas inversiones. Al representar las sustituciones en forma de tablas hemos convenido en que los elementos del conjunto que figuran en la fila superior pueden ser dispuestos en un orden arbitrario. Sin embargo, en el caso en que el conjunto considerado es la colección de los números enteros, podemos convenir en escribir los números de la primera fila en el orden de crecimiento. En este caso la sustitución quedará plenamente determinada al indicar solamente la fila inferior de la tabla, es decir, al indicar *una permutación* de números. La correspondencia entre las sustituciones y las permutaciones, obtenida de esta forma, resulta biyectiva y obtenemos la posibilidad de representar las sustituciones en forma de tablas de una fila en lugar de tablas de dos filas. Sea σ una sustitución que corresponde a la permutación $i_1, i_2, \ldots, i_n$ de los números $1, 2, \ldots, n$. Consideremos todos los pares i_k, i_l $(k < l)$; diremos que el par i_k, i_l forma *una inversión*, si $i_k > i_l$. De la relación

$$\Delta\sigma = \prod_{k<l} (x_{k\sigma} - x_{l\sigma}) = \prod (x_{i_k} - x_{i_l})$$

se ve que $\varepsilon_\sigma = 1$, si es par el número de factores $x_{i_k} - x_{i_l}$ para los cuales $i_k > i_l$, y que $\varepsilon_\sigma = -1$, si el número de estos factores es impar. Por consiguiente, la paridad de la sustitución σ coincide con la paridad del número de inversiones en la permutación $i_1, i_2, \ldots, i_n$.

Ejemplos y problemas

1. Tomemos en el espacio corriente un sistema rectangular de coordenadas $OXYZ$. Sea $\mathcal{A}$ el giro del espacio alrededor del eje OX en 90° en dirección de OY hacia OZ, sea $\mathcal{B}$ el giro en 90° alrededor del eje OY en dirección de OZ hacia OX y sea $\mathcal{C}$ el giro en 90° alrededor del eje OZ en dirección de OX hacia OY. Sea m el punto de coordenadas (1, 0, 1). Calcúlense las coordenadas de los puntos $m\mathcal{A}$, $m(\mathcal{BC})$ y $m(\mathcal{CB})$. Demuéstrese que

$$\mathcal{A}^4 = \mathcal{B}^4 = \mathcal{C}^4 = \mathcal{E}, \quad \mathcal{AB} \neq \mathcal{BA} \text{ y } \mathcal{A}^2\mathcal{B}^2 = \mathcal{B}^2\mathcal{A}^2.$$

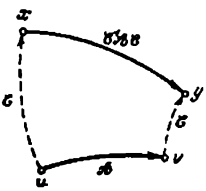


Fig. 3.

2. Consideremos dos aplicaciones $\mathcal{A}$ y $\mathcal{C}$ de un conjunto arbitrario $\mathfrak{M}$ aceptando que $\mathcal{C}$ es invertible. Tomemos un elemento cualquiera u de $\mathfrak{M}$ y sea $v = u\mathcal{A}$. La aplicación $\mathcal{C}$ «traslada» los elementos u y v en unos elementos x e y. Demuéstrese que en estas condiciones la aplicación $\mathcal{C}^{-1}\mathcal{AC}$ transforma x en y (fig. 3).

3. Sea O un punto arbitrario del plano, sea $\mathcal{A}$ el giro del plano alrededor del punto O en un ángulo determinado α y sea $\mathcal{B}$ el traslado paralelo del plano a una distancia a en una dirección determinada. Demuéstrese que $\mathcal{B}^{-1}\mathcal{AB}$ es el giro en el ángulo α alrededor del punto $O\mathcal{B}$ y que $\mathcal{A}^{-1}\mathcal{BA}$ es el traslado del plano a la distancia a en la dirección que se obtiene girando en el ángulo α la dirección inicial.

§ 8. Aplicaciones lineales y sus matrices

8.1. Propiedades elementales. Se llama *aplicación* de un espacio lineal $\mathfrak{L}$ una ley que a todo vector de $\mathfrak{L}$ pone en correspondencia de nuevo un vector determinado de $\mathfrak{L}$. Una aplicación se llama *lineal*, si transforma el producto de un número por un vector en el producto del mismo número por el vector correspondiente y la suma de vectores en la suma de los vectores correspondientes. Abreviando, una aplicación $\mathcal{A}$ se llama lineal, si para cualesquiera vectores x e y del espacio $\mathfrak{L}$ y para cualquier número α del cuerpo de coeficientes tienen lugar las igualdades

$$(\alpha x)\mathcal{A} = \alpha(x\mathcal{A}), \tag{1}$$

$$(x+y)\mathcal{A} = x\mathcal{A} + y\mathcal{A}. \tag{2}$$

Tomando $\alpha = 0$ en (1), obtenemos

$$o\mathcal{A} = o,$$

es decir, *toda aplicación lineal transforma el vector nulo en el vector nulo.*

De (1) y (2) también se desprende directamente la siguiente propiedad principal de las aplicaciones lineales: *si $\mathcal{A}$ es una aplicación lineal, se tiene*

$$(\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_m x_m)\mathcal{A} = \alpha_1(x_1\mathcal{A}) + \alpha_2(x_2\mathcal{A}) + \ldots + \alpha_m(x_m\mathcal{A}), \tag{3}$$

donde $x_1, x_2, \ldots, x_m$ *son vectores arbitrarios de* $\mathfrak{L}$ y $\alpha_1, \alpha_2, \ldots, \alpha_m$ *son números arbitrarios del cuerpo de coeficientes.*

Para la demostración basta aplicar la inducción según m. Para $m=1$ la fórmula (3) coincide con (1). Supongamos ahora que (3) es válida en el caso de $m-1$ sumandos. Entonces

$$(\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_m x_m)\mathcal{A} = (\alpha_1 x_1 + (\alpha_2 x_2 + \ldots + \alpha_m x_m))\mathcal{A} =$$
$$= (\alpha_1 x_1)\mathcal{A} + (\alpha_2 x_2 + \ldots + \alpha_m x_m)\mathcal{A} =$$
$$= \alpha_1 (x_1 \mathcal{A}) + \alpha_2 (x_2 \mathcal{A}) + \ldots + \alpha_m (x_m \mathcal{A}),$$

que es lo que se quería demostrar. Para $m=2$, la fórmula (3) se convierte en

$$(\alpha x + \beta y)\mathcal{A} = \alpha (x\mathcal{A}) + \beta (y\mathcal{A}), \qquad (4)$$

de donde para $\beta = 0$ y $\alpha = \beta = 1$ se obtiene de nuevo (1) y (2). Por consiguiente, la propiedad (4) caracteriza totalmente las aplicaciones lineales y puede servir de definición de las mismas.

Hemos convenido anteriormente en llamar aplicación idéntica $\mathcal{E}$ aquella aplicación que transforma todo elemento en sí mismo. Por esto, si el conjunto considerado es la colección de los vectores de un espacio lineal $\mathfrak{L}$, se tiene

$$(\alpha x + \beta y)\mathcal{E} = \alpha x + \beta y = \alpha (x\mathcal{E}) + \beta (y\mathcal{E}).$$

Por consiguiente, *la aplicación idéntica de un espacio vectorial es lineal.* La aplicación que transforma todo vector en el vector nulo se llama *nula* y se indica por $\mathcal{O}$. Está claro que la aplicación nula también es lineal.

Consideremos dos ejemplos concretos. Sea $\mathfrak{R}$ el espacio vectorial corriente, es decir, el conjunto de segmentos orientados que parten de un punto fijo O. Consideremos un plano que pasa por O e indiquemos mediante $x\mathcal{A}$ la proyección del segmento x sobre este plano. Entonces las conocidas propiedades de las proyecciones — 1) la proyección de una suma de segmentos es igual a la suma de sus proyecciones y 2) si un segmento es aumentado en α veces, su proyección también se aumenta en α veces — pueden ser representadas en forma de las igualdades

$$(x+y)\mathcal{A} = x\mathcal{A} + y\mathcal{A} \quad \text{y} \quad (\alpha x)\mathcal{A} = \alpha (x\mathcal{A}),$$

de las cuales se desprende que la operación de proyección es una aplicación lineal.

Como segundo ejemplo consideremos el conjunto de todos los polinomios en la variable λ de orden no mayor que n. Estos polinomios forman, respecto a las operaciones corrientes de adición y de multiplicación por número, un espacio lineal de dimensión $n+1$. Pongamos en correspondencia a todo polinomio su derivada. Puesto que la derivada de una suma es igual a la suma de las derivadas

de los sumandos y puesto que un factor constante puede ser extraído del signo de la derivada, resulta que la operación de diferenciación es una aplicación lineal del espacio de polinomios.

Veamos ahora mediante qué elementos puede ser definida una aplicación lineal.

TEOREMA *Sea* $a_1, a_2, \ldots, a_n$ *una base de un espacio lineal* $\mathfrak{L}$. *Tomemos en* $\mathfrak{L}$ *unos vectores absolutamente arbitrarios* $b_1, b_2, \ldots, b_n$. *Entonces existe una aplicación lineal del espacio* $\mathfrak{L}$, *y sólo una, que transforma los vectores* $a_1, a_2, \ldots, a_n$ *en los vectores* $b_1, b_2, \ldots, b_n$ *respectivamente.*

Para construir la aplicación deseada, tomemos un vector arbitrario x y representémoslo linealmente en términos de la base $a_1, a_2, \ldots, a_n$. Sea

$$x = \xi_1 a_1 + \xi_2 a_2 + \ldots + \xi_n a_n. \tag{5}$$

Consideremos el vector

$$x' = \xi_1 b_1 + \xi_2 b_2 + \ldots + \xi_n b_n.$$

Indiquemos por $\mathcal{A}$ la aplicación que transforma x en x'. Por consiguiente, si x está representado por (5), se tiene

$$x\mathcal{A} = \xi_1 b_1 + \xi_2 b_2 + \ldots + \xi_n b_n. \tag{6}$$

Tomando aquí $\xi_j = 1$ y $\xi_j = 0$ $(j \neq i,\ j = 1, 2, \ldots, n)$, obtenemos $a_i\mathcal{A} = b_i$ $(i = 1, 2, \ldots, n)$, de modo que la aplicación $\mathcal{A}$ transforma los vectores $a_1, \ldots, a_n$ en $b_1, \ldots, b_n$. Demostremos que $\mathcal{A}$ es *lineal*. Multiplicando (5) por un número cualquiera α, tendremos

$$\alpha x = (\alpha\xi_1) a_1 + (\alpha\xi_2) a_2 + \ldots + (\alpha\xi_n) a_n.$$

Comparando este resultado con (6), obtenemos

$$(\alpha x)\mathcal{A} = (\alpha\xi_1) b_1 + (\alpha\xi_2) b_2 + \ldots + (\alpha\xi_n) b_n,$$

es decir,

$$(\alpha x)\mathcal{A} = \alpha (x\mathcal{A}). \tag{7}$$

Sea

$$y = \eta_1 a_1 + \eta_2 a_2 + \ldots + \eta_n a_n$$

otro vector de $\mathfrak{L}$. Entonces se tiene

$$x + y = (\xi_1 + \eta_1) a_1 + (\xi_2 + \eta_2) a_2 + \ldots + (\xi_n + \eta_n) a_n$$

y, por consiguiente,

$$y\mathcal{A} = \eta_1 b_1 + \eta_2 b_2 + \ldots + \eta_n b_n,$$
$$(x + y)\mathcal{A} = (\xi_2 + \eta_1) b_1 + \ldots + (\xi_n + \eta_n) b_n.$$

De la última igualdad resulta

$$(x + y)\mathcal{A} = \xi_1 b_1 + \ldots + \xi_n b_n + \eta_1 b_1 + \ldots + \eta_n b_n = x\mathcal{A} + y\mathcal{A}. \tag{8}$$

Las propiedades (7) y (8) significan que $\mathcal{A}$ es lineal.

Resta probar que toda aplicación lineal $\mathcal{B}$ que transforma $a_1, \ldots, a_n$ en $b_1, \ldots, b_n$ coincide con $\mathcal{A}$. Por hipótesis, $a_i\mathcal{B} = b_i$ $(i = 1, 2, \ldots, n)$. Luego, si el vector x está representado por (5), se tiene

$$x\mathcal{B} = \xi_1 (a_1\mathcal{B}) + \ldots + \xi_n (a_n\mathcal{B}) = \xi_1 b_1 + \ldots + \xi_n b_n = x\mathcal{A},$$

es decir, $\mathcal{B} = \mathcal{A}$. Hemos demostrado el teorema.

8.2. Matriz de una aplicación lineal. Veremos ahora cómo se puede definir una aplicación lineal mediante números. Tomemos en el espacio $\mathfrak{L}$ un sistema cualquiera de coordenadas $a_1, a_2, \ldots, a_n$ y supongamos dada una aplicación lineal $\mathcal{A}$ de este espacio. La aplicación $\mathcal{A}$ transforma los vectores $a_1, \ldots, a_n$ en unos vectores $a_1\mathcal{A}, a_2\mathcal{A}, \ldots, a_n\mathcal{A}$ que pueden ser expresados linealmente en términos de $a_1, a_2, \ldots, a_n$. Sean

$$\begin{aligned}
a_1\mathcal{A} &= \alpha_{11}a_1 + \alpha_{12}a_2 + \ldots + \alpha_{1n}a_n,\\
a_2\mathcal{A} &= \alpha_{21}a_1 + \alpha_{22}a_2 + \ldots + \alpha_{2n}a_n,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\\
a_n\mathcal{A} &= \alpha_{n1}a_1 + \alpha_{n2}a_2 + \ldots + \alpha_{nn}a_n
\end{aligned}$$

estas expresiones. La matriz

$$A = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \ldots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \ldots & \alpha_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ \alpha_{n1} & \alpha_{n2} & \ldots & \alpha_{nn} \end{bmatrix},$$

formada por las filas coordenadas de los vectores $a_1\mathcal{A}, \ldots, a_n\mathcal{A}$, se llama *matriz de la aplicación* $\mathcal{A}$ en el sistema de coordenadas $a_1, a_2, \ldots, a_n$. Por consiguiente, dado un sistema de coordenadas, a toda aplicación lineal corresponde una matriz determinada. Surge la pregunta de si esta correspondencia es biyectiva. La respuesta es positiva ya que conociendo la matriz A podemos encontrar primero los vectores

$$\begin{aligned}
b_1 &= \alpha_{11}a_1 + \alpha_{12}a_2 + \ldots + \alpha_{1n}a_n,\\
b_2 &= \alpha_{21}a_1 + \alpha_{22}a_2 + \ldots + \alpha_{2n}a_n,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\\
b_n &= \alpha_{n1}a_1 + \alpha_{n2}a_2 + \ldots + \alpha_{nn}a_n
\end{aligned}$$

y después construir, de acuerdo con el teorema del punto anterior, la aplicación lineal $\mathcal{A}$ que transforma $a_1, \ldots, a_n$ en $b_1, \ldots, b_n$ respectivamente. Esta aplicación es única y su matriz coincide, obviamente, con la matriz A que es lo que se quería demostrar.

Veamos qué matrices corresponden a las aplicaciones idéntica y nula del espacio $\mathfrak{L}$. Sea $a_1, a_2, \ldots, a_n$ un sistema arbitrario de

coordenadas de $\mathfrak{L}$. Entonces tendremos

$$a_1\mathcal{O} = 0 \cdot a_1 + \ldots + 0 \cdot a_n, \qquad a_1\mathcal{E} = 1 \cdot a_1 + \ldots + 0 \cdot a_n,$$
$$\ldots\ldots\ldots\ldots \quad \text{y} \quad \ldots\ldots\ldots\ldots$$
$$a_n\mathcal{O} = 0 \cdot a_1 + \ldots + 0 \cdot a_n \qquad a_n\mathcal{E} = 0 \cdot a_1 + \ldots + 1 \cdot a_n.$$

Por consiguiente, *la aplicación nula tiene la matriz nula y la aplicación idéntica tiene la matriz unidad.*

Planteémonos el problema: ¿cómo conociendo la matriz de una aplicación lineal $\mathcal{A}$ y las coordenadas de un vector x determinar las coordenadas del vector $x\mathcal{A}$?

Sea $a_1, a_2, \ldots, a_n$ el sistema de coordenadas escogido y sean $\xi_1, \xi_2, \ldots, \xi_n$ las coordenadas del vector x en el mismo. Indiquemos por α_{ij} $(i, j = 1, \ldots, n)$ los elementos de la matriz A de la aplicación $\mathcal{A}$ calculados en este mismo sistema de coordenadas. Tenemos

$$x = \xi_1 a_1 + \xi_2 a_2 + \ldots + \xi_n a_n,$$
$$x\mathcal{A} = \xi_1 (a_1\mathcal{A}) + \xi_2 (a_2\mathcal{A}) + \ldots + \xi_n (a_n\mathcal{A}).$$

Según la definición de la matriz de una aplicación se tiene

$$a_i\mathcal{A} = \alpha_{i1}a_1 + \alpha_{i2}a_2 + \ldots + \alpha_{in}a_n \quad (i = 1, \ldots, n).$$

Introduciendo estos valores en la igualdad anterior, obtenemos

$$x\mathcal{A} = (\xi_1\alpha_{11} + \xi_2\alpha_{21} + \ldots + \xi_n\alpha_{n1})\, a_1 + \ldots$$
$$\ldots + (\xi_1\alpha_{1n} + \xi_2\alpha_{2n} + \ldots + \xi_n\alpha_{nn})\, a_n.$$

Por consiguiente, la fila de coordenadas del nuevo vector $x\mathcal{A}$ es

$$[x\mathcal{A}] = [\xi_1\alpha_{11} + \ldots + \xi_n\alpha_{n1}, \ldots, \xi_1\alpha_{1n} + \ldots + \xi_n\alpha_{nn}] = [x]\, A,$$

es decir, *la fila de coordenadas del vector nuevo es igual a la fila de coordenadas del vector antiguo multiplicada por la matriz de la aplicación lineal:*

$$[x\mathcal{A}] = [x]\, A.$$

8.3. Transformación de coordenadas. En el punto anterior ha sido establecida una correspondencia biyectiva entre las aplicaciones lineales de un espacio vectorial $\mathfrak{L}$ de n dimensiones y las matrices cuadradas de orden n. Sin embargo, para ello ha sido necesario escoger primero en $\mathfrak{L}$ un determinado sistema de coordenadas. Si cambiamos éste, cambiamos la correspondencia. Tendremos como resultado que a una misma aplicación lineal $\mathcal{A}$ corresponderán en los sistemas antiguo y nuevo de coordenadas diferentes matrices A y A_1. Hallemos la relación entre las mismas.

Sean $a_1, a_2, \ldots, a_n$ y $a'_1, a'_2, \ldots, a'_n$ el antiguo y el nuevo sistemas de coordenadas y sea T la matriz del cambio (véase el p. 5.1). Indiquemos por $[x]$ y $[x]_1$ las filas de coordenadas del

vector x en el antiguo y en el nuevo sistemas de coordenadas. De acuerdo con la regla de una aplicación lineal, tenemos en el sistema antiguo de coordenadas

$$[x\mathcal{A}] = [x]\,A.$$

En el nuevo sistema de coordenadas esta misma regla da

$$[x\mathcal{A}]_1 = [x]_1 A_1.$$

Sin embargo, la regla de transformación de coordenadas del p. 5.1 muestra que

$$[x] = [x]_1 T \quad \text{y} \quad [x\mathcal{A}] = [x\mathcal{A}]_1 T.$$

Introduciendo aquí los valores de $[x\mathcal{A}]$, $[x\mathcal{A}]_1$ y $[x]$ de las igualdades anteriores, obtenemos

$$[x]_1 TA = [x]_1 A_1 T.$$

Puesto que el vector x es arbitrario, tenemos, aplicando el lema del p. 5.1, $TA = A_1 T$, o

$$A_1 = TAT^{-1}.$$

Es decir, *la matriz de una aplicación lineal en el sistema nuevo de coordenadas es igual a la matriz de esta misma aplicación lineal en el sistema antiguo transformada por la matriz del cambio recíproco.*

Para concluir consideremos el problema de cómo puede ser interpretada, desde el punto de vista de la teoría de aplicaciones lineales, la matriz del cambio de un sistema de coordenadas por otro. Sean $a_1, a_2, \ldots, a_n$ y $a'_1, a'_2, \ldots, a'_n$ dos sistemas de coordenadas dados de un espacio $\mathfrak{L}$. Según el teorema del p. 8.1, existe una aplicación lineal $\mathcal{T}$, determinada unívocamente, que transforma el sistema antiguo de vectores coordenados en el sistema nuevo, es decir, que posee la propiedad

$$a_i\mathcal{T} = a'_i \quad (i = 1, 2, \ldots, n).$$

Para escribir ahora, según la regla del p. 8.2, la matriz de la aplicación $\mathcal{T}$ en el sistema antiguo, tenemos que expresar todos los vectores a'_i en términos de $a_1, \ldots, a_n$. Pero lo mismo tenemos que hacer para obtener la matriz del cambio. Por consiguiente, *la matriz del cambio es la matriz de la aplicación lineal que transforma el sistema antiguo de coordenadas en el sistema nuevo*, calculada en el sistema antiguo de coordenadas. Es más, la última indicación es innecesaria, ya que al calcular la matriz de la aplicación $\mathcal{T}$ en el sistema nuevo de coordenadas obtendremos la misma matriz del cambio T. En efecto, si la matriz de la aplicación $\mathcal{T}$ en el sistema antiguo de coordenadas es T, la matriz del cambio también será T. Por lo tanto, la matriz de $\mathcal{T}$ en el nuevo sistema será igual a $TTT^{-1} = T$, que es lo que se quería demostrar.

Ejemplos y problemas

1. Consideremos el espacio de los vectores que pertenecen a un plano y parten de un punto O. Demuéstrese que la aplicación, consistente en el giro de todos los vectores en un ángulo α alrededor del punto O, es lineal y que su matriz es igual a $\begin{bmatrix} \cos\alpha & \operatorname{sen}\alpha \\ -\operatorname{sen}\alpha & \cos\alpha \end{bmatrix}$, siempre que el sistema de coordenadas esté formado por dos vectores perpendiculares de longitud 1.

2. Sea $\mathfrak{R}$ el espacio corriente de los segmentos orientados que parten de un punto O. Tomemos en $\mathfrak{R}$ un sistema de coordenadas formado por tres vectores perpendiculares e_1, e_2 y e_3 de longitud 1. Demuéstrese que la matriz de la aplicación lineal $\mathcal{P}$, consistente en la proyección de los vectores sobre el eje e_1, es igual a $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, mientras que la matriz de la proyección sobre el plano e_1e_2 es igual a $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

3. ¿Cómo cambia la matriz de una aplicación lineal $\mathcal{A}$, si en el sistema de coordenadas $e_1, e_2, \ldots, e_n$ se cambian entre sí dos vectores cualesquiera, por ejemplo, e_1 y e_2?

4. Toda aplicación lineal de un espacio de dimensión uno consiste en la multiplicación de todos sus vectores por un mismo número.

5. Sea $\mathfrak{L}$ el espacio de todas las matrices cuadradas de segundo orden. Demuéstrese que la aplicación $\mathcal{A}$, consistente en la multiplicación a la derecha de todas las matrices de $\mathfrak{L}$ por la matriz $\begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix}$, es lineal. Hállese la matriz de la aplicación $\mathcal{A}$ si como sistema de coordenadas de $\mathfrak{L}$ se toma el sistema

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \text{ y } \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

6. Una aplicación lineal $\mathcal{A}$ de un espacio $\mathfrak{L}$ de cuatro dimensiones tiene en el sistema de coordenadas e_1, e_2, e_3 y e_4 la matriz

$$\begin{bmatrix} 1 & 2 & 3 & 2 \\ -1 & 0 & 3 & 1 \\ 2 & 1 & 5 & -1 \\ 1 & 1 & 2 & 2 \end{bmatrix}.$$

¿Cuál será la matriz de esta aplicación, si para el nuevo sistema de coordenadas se toma el sistema 1) e_1, e_2, e_3 y e_4; 2) $e_1, e_1+e_2, e_1+e_2+e_3$ y $e_1+e_2+e_3+e_4$?

§ 9. Operaciones con aplicaciones lineales

9.1. Multiplicación de aplicaciones lineales. Sean $\mathcal{A}$ y $\mathcal{B}$ dos aplicaciones lineales definidas en un espacio lineal $\mathfrak{L}$. Aplicando a cualquier vector x de $\mathfrak{L}$ primero la aplicación $\mathcal{A}$ y después la aplicación $\mathcal{B}$, obtendremos un vector

$$y = (x\mathcal{A})\mathcal{B}.$$

La aplicación que transforma x directamente en y ha sido llamada en el p. 7.2 producto de $\mathcal{A}$ por $\mathcal{B}$. Por consiguiente,

$$x(\mathcal{A}\mathcal{B}) = (x\mathcal{A})\mathcal{B}. \tag{1}$$

Demostremos que *el producto de dos aplicaciones lineales es una aplicación lineal*. Para ello es suficiente demostrar, según el p. 8.1, que tiene lugar la igualdad

$$(\alpha x + \beta y)(\mathcal{AB}) = \alpha \cdot x(\mathcal{AB}) + \beta \cdot y(\mathcal{AB}).$$

En virtud de (1) se tiene

$$(\alpha x + \beta y)(\mathcal{AB}) = ((\alpha x + \beta y)\mathcal{A})\mathcal{B}.$$

Puesto que $\mathcal{A}$ y $\mathcal{B}$ son lineales, resulta

$$((\alpha x + \beta y)\mathcal{A})\mathcal{B} = (\alpha(x\mathcal{A}) + \beta(y\mathcal{A}))\mathcal{B} = \alpha \cdot (x\mathcal{A})\mathcal{B} + \beta \cdot (y\mathcal{A})\mathcal{B},$$

es decir,

$$(\alpha x + \beta y)(\mathcal{AB}) = \alpha \cdot x(\mathcal{AB}) + \beta \cdot y(\mathcal{AB})$$

que es lo que se quería demostrar.

Tomemos en el espacio $\mathfrak{L}$ un sistema de coordenadas e indiquemos por A y B las matrices de las aplicaciones $\mathcal{A}$ y $\mathcal{B}$. ¿Cómo hallar la matriz de la aplicación $\mathcal{AB}$? Indiquemos esta matriz incógnita por C. Sea x un vector cualquiera del espacio $\mathfrak{L}$ y sea $[x]$ su fila de coordenadas. Según la regla de aplicación lineal, se tiene

$$[x(\mathcal{AB})] = [x]C.$$

Por otro lado, tenemos

$$[x(\mathcal{AB})] = [(x\mathcal{A})\mathcal{B}] = [x\mathcal{A}]B = [x]AB.$$

Comparando ambos resultados vemos que

$$[x]C = [x]AB$$

cualquiera que sea x. De acuerdo con el lema del p. 5.1 de aquí se deduce que

$$C = AB,$$

es decir, *la matriz de un producto de aplicaciones lineales es igual al producto de las matrices de estas aplicaciones*.

Consideremos una aplicación lineal *biyectiva* $\mathcal{A}$. Si $\mathcal{A}$ transforma el vector x en y, la aplicación que transforma y en x será la aplicación *inversa* $\mathcal{A}^{-1}$ (véase el p. 7.3). Demostremos que *siendo* $\mathcal{A}$ *lineal, también* $\mathcal{A}^{-1}$ *es lineal*. En efecto, sea

$$u\mathcal{A}^{-1} = x \quad \text{y} \quad v\mathcal{A}^{-1} = y,$$

donde u y v son vectores arbitrarios de $\mathfrak{L}$. Puesto que la aplicación $\mathcal{A}$ es lineal, se tiene

$$(\alpha x + \beta y)\mathcal{A} = \alpha \cdot x\mathcal{A} + \beta \cdot y\mathcal{A} = \alpha u + \beta v,$$

de donde resulta

$$(\alpha u + \beta v)\mathcal{A}^{-1} = \alpha x + \beta y = \alpha \cdot u\mathcal{A}^{-1} + \beta \cdot v\mathcal{A}^{-1}$$

que es lo que se quería demostrar.

Siendo A la matriz de la aplicación $\mathcal{A}$ y X la matriz de la aplicación $\mathcal{A}^{-1}$, de las relaciones

$$\mathcal{A}\mathcal{A}^{-1} = \mathcal{A}^{-1}\mathcal{A} = \mathcal{E}$$

se deduce que

$$AX = XA = E,$$

es decir, $X = A^{-1}$. Por consiguiente, *la aplicación inversa tiene la matriz inversa.* En particular, *para que una aplicación lineal sea invertible es necesario y suficiente que sea invertible su matriz.*

De nuestros resultados también se desprende directamente la siguiente regla: *la matriz de la aplicación $\mathcal{A}^n$ es igual a A^n, donde A es la matriz de la aplicación $\mathcal{A}$.*

9.2. Multiplicación por número y adición. En este punto y en el que sigue se supone que el cuerpo principal K es un cuerpo conmutativo. La aplicación que resulta al aplicar a un vector primero la aplicación $\mathcal{A}$ y al multiplicar después el vector nuevo por un número α se llama *producto* de α por $\mathcal{A}$ y se indica por $\alpha\mathcal{A}$. Esta definición se puede expresar mediante la fórmula

$$x(\alpha\mathcal{A}) = \alpha(x\mathcal{A}).$$

Razonando igual que en el caso de la multiplicación de aplicaciones lineales, veremos fácilmente que *siendo $\mathcal{A}$ una aplicación lineal, también $\alpha\mathcal{A}$ es una aplicación lineal.* Determinemos su matriz. Supongamos que en un sistema de coordenadas fijado la aplicación $\mathcal{A}$ tiene la matriz A y la aplicación $\alpha\mathcal{A}$ tiene la matriz B. Según la regla de una aplicación lineal, tenemos

$$[x(\alpha\mathcal{A})] = [x]\,B \quad \text{y}$$
$$[x(\alpha\mathcal{A})] = [\alpha(x\mathcal{A})] = \alpha[x\mathcal{A}] = \alpha[x]\,A = [x](\alpha A),$$

de donde resulta que $[x]\,B = [x](\alpha A)$, es decir, $B = \alpha A$.

Luego, *la matriz del producto de un número por una aplicación lineal es igual al producto de este número por la matriz de la aplicación.*

Indiquemos, finalmente, las fórmulas

$$\alpha(\beta\mathcal{A}) = (\alpha\beta)\,\mathcal{A},$$
$$0\cdot\mathcal{A} = \mathcal{O},$$
$$1\cdot\mathcal{A} = \mathcal{A},$$
$$\alpha(\mathcal{A}\mathcal{B}) = (\alpha\mathcal{A})\,\mathcal{B} = \mathcal{A}(\alpha\mathcal{B}),$$

análogas plenamente a las fórmulas correspondientes del cálculo de matrices. La demostración queda a cargo del lector.

Tomemos ahora dos aplicaciones $\mathcal{A}$ y $\mathcal{B}$ de un espacio lineal $\mathfrak{L}$ y pongamos en correspondencia a todo vector x de $\mathfrak{L}$ el vector $x\mathcal{A} + x\mathcal{B}$. La aplicación que transforma x en $x\mathcal{A} + x\mathcal{B}$ se indica

por $\mathcal{A}+\mathcal{B}$ y se denomina *suma* de $\mathcal{A}$ y $\mathcal{B}$. Es decir, por definición,

$$x(\mathcal{A}+\mathcal{B}) = x\mathcal{A}+x\mathcal{B}.$$

Si las aplicaciones $\mathcal{A}$ y $\mathcal{B}$ son lineales, se tiene

$$(\alpha x+\beta y)(\mathcal{A}+\mathcal{B}) = (\alpha x+\beta y)\mathcal{A}+(\alpha x+\beta y)\mathcal{B} =$$
$$= \alpha\cdot x\mathcal{A}+\beta\cdot y\mathcal{A}+\alpha\cdot x\mathcal{B}+\beta\cdot y\mathcal{B} =$$
$$= \alpha(x\mathcal{A}+x\mathcal{B})+\beta(y\mathcal{A}+y\mathcal{B}) = \alpha\cdot x(\mathcal{A}+\mathcal{B})+\beta\cdot y(\mathcal{A}+\mathcal{B}).$$

Por consiguiente, *la suma de dos aplicaciones lineales es una aplicación lineal.* Determinemos su matriz. Sean A y B las matrices de las aplicaciones $\mathcal{A}$ y $\mathcal{B}$ en un sistema de coordenadas. Sea C la matriz de la aplicación $\mathcal{A}+\mathcal{B}$. Tenemos

$$[x(\mathcal{A}+\mathcal{B})] = [x]C,$$

$$[x(\mathcal{A}+\mathcal{B})] = [x\mathcal{A}+x\mathcal{B}] = [x\mathcal{A}]+[x\mathcal{B}] = [x]A+[x]B = [x](A+B).$$

Por consiguiente,

$$[x]C = [x](A+B) \quad \text{y} \quad C = A+B,$$

es decir, *la matriz de una suma de aplicaciones es igual a la suma de sus matrices.*

Las operaciones con aplicaciones lineales se rigen por las mismas leyes que las operaciones con matrices. Una parte de ellas, las relacionadas solamente con la multiplicación, ya las hemos indicado. Señalemos ahora también aquellas que se refieren o bien a la adición o bien a las relaciones entre la adición y multiplicación:

$$\mathcal{A}+\mathcal{B} = \mathcal{B}+\mathcal{A},$$
$$(\mathcal{A}+\mathcal{B})+\mathcal{C} = \mathcal{A}+(\mathcal{B}+\mathcal{C}),$$
$$\mathcal{A}+\mathcal{O} = \mathcal{A},$$
$$\alpha(\mathcal{A}+\mathcal{B}) = \alpha\mathcal{A}+\alpha\mathcal{B},$$
$$(\alpha+\beta)\mathcal{A} = \alpha\mathcal{A}+\beta\mathcal{A},$$
$$\mathcal{A}(\mathcal{B}+\mathcal{C}) = \mathcal{A}\mathcal{B}+\mathcal{A}\mathcal{C},$$
$$(\mathcal{A}+\mathcal{B})\mathcal{C} = \mathcal{A}\mathcal{C}+\mathcal{B}\mathcal{C}.$$

Todas estas igualdades se demuestran siguiendo un mismo método: se toma un vector arbitrario x y se demuestra que la aplicación que figura en el primer miembro transforma x en el mismo vector en el que lo transforma la aplicación que figura en el segundo miembro. Por ejemplo,

$$x\mathcal{A}(\mathcal{B}+\mathcal{C}) = (x\mathcal{A})(\mathcal{B}+\mathcal{C}) = (x\mathcal{A})\mathcal{B}+(x\mathcal{A})\mathcal{C},$$
$$x(\mathcal{A}\mathcal{B}+\mathcal{A}\mathcal{C}) = x(\mathcal{A}\mathcal{B})+x(\mathcal{A}\mathcal{C}) = (x\mathcal{A})\mathcal{B}+(x\mathcal{A})\mathcal{C},$$

de donde resulta que $\mathcal{A}(\mathcal{B}+\mathcal{C}) = \mathcal{A}\mathcal{B}+\mathcal{A}\mathcal{C}$.

Todas estas leyes también se pueden deducir directamente de las fórmulas correspondientes del cálculo de matrices. En efecto, existe una correspondencia biyectiva entre las matrices cuadradas de orden n y las aplicaciones lineales de un espacio lineal de n dimensiones. Esta correspondencia posee la propiedad de transformar la suma en la suma y el producto en el producto. Por esto, toda identidad entre matrices implica una identidad análoga entre aplicaciones lineales, que es lo que se quería demostrar.

9.3. Polinomios en aplicaciones lineales. Sea

$$f(\lambda) = \alpha_0 + \alpha_1\lambda + \ldots + \alpha_m\lambda^m$$

un polinomio en la variable λ. La expresión

$$f(\mathcal{A}) = \alpha_0\mathcal{E} + \alpha_1\mathcal{A} + \ldots + \alpha_m\mathcal{A}^m,$$

donde $\mathcal{E}$ es la aplicación idéntica y $\mathcal{A}$ es una aplicación cualquiera, se llama valor del polinomio $f(\lambda)$ para $\lambda = \mathcal{A}$ o simplemente *polinomio en* $\mathcal{A}$. *Si la matriz de la aplicación* $\mathcal{A}$ *en un sistema de coordenadas es igual a* A, *la matriz de la aplicación* $f(\mathcal{A})$ *en este mismo sistema de coordenadas es*

$$f(A) = \alpha_0 E + \alpha_1 A + \ldots + \alpha_m A^m.$$

Efectivamente, $f(\mathcal{A})$ se obtiene de $\mathcal{A}$ mediante las operaciones de multiplicación, de multiplicación por número y de adición. De los resultados señalados anteriormente se ve que realizando estas mismas operaciones con la matriz A se obtiene la matriz de la aplicación $f(\mathcal{A})$.

Todas las reglas de operaciones con polinomios en una variable tienen lugar también para los polinomios en una aplicación lineal. Por esto, si en alguna identidad entre polinomios en λ se sustituye λ por una aplicación lineal, se obtiene una relación verídica. Por ejemplo, de las identidades

$$\lambda^2 - 1 = (\lambda - 1)(\lambda + 1) \quad \text{y} \quad (\lambda + 1)^2 + (\lambda - 1)^2 - 2\lambda^2 = 2$$

se obtiene realizando la sustitución $\lambda = \mathcal{A}$

$$\mathcal{A}^2 - \mathcal{E} = (\mathcal{A} - \mathcal{E})(\mathcal{A} + \mathcal{E}) \text{ y}$$
$$(\mathcal{A} + \mathcal{E})^2 + (\mathcal{A} - \mathcal{E})^2 - 2\mathcal{A}^2 = 2\mathcal{E}.$$

En particular, la identidad

$$f(\lambda)\,g(\lambda) = g(\lambda)\,f(\lambda)$$

implica la relación

$$f(\mathcal{A})\,g(\mathcal{A}) = g(\mathcal{A})\,f(\mathcal{A})$$

que significa que los polinomios en una misma aplicación lineal siempre conmutan.

La situación resulta diferente en el caso de polinomios en varias variables. Esto se debe a que en los cálculos con los polinomios se

acepta que las variables conmutan. Por esto tenemos, por ejemplo, las igualdades $\lambda\mu\lambda = \lambda^2\mu$, $\lambda\mu^2\lambda\mu = \lambda^2\mu^2$, etc.

Sustituyendo en estas igualdades las variables λ y μ por unas aplicaciones lineales arbitrarias $\mathcal{A}$ y $\mathcal{B}$, obtenemos las relaciones

$$\mathcal{ABA} = \mathcal{A}^2\mathcal{B} \quad \text{y} \quad \mathcal{AB}^2\mathcal{AB} = \mathcal{A}^2\mathcal{B}^2$$

que pueden resultar falsas para determinadas aplicaciones lineales. Está claro que estas dificultades desaparecen, si las aplicaciones consideradas conmutan. Por consiguiente, en toda identidad entre polinomios en varias incógnitas se pueden sustituir estas incógnitas por aplicaciones lineales arbitrarias que conmutan, obteniéndose como resultado una relación verídica entre aplicaciones lineales.

Hemos convenido en que todos los espacios que se consideran en este libro son de dimensión finita. Sin embargo, esto no excluye el hecho de que una parte de definiciones y de teoremas tenga lugar también para los espacios de dimensión infinita. A título de ejemplo, podemos señalar las definiciones de aplicaciones lineales y de las operaciones con las mismas y aquellas propiedades de las aplicaciones lineales, expuestas en este parágrafo, que no están relacionadas con matrices.

Ejemplos y problemas

1. Sea $\mathfrak{L}$ el conjunto de todos los polinomios en λ de grado $\leqslant n$. Sea $\mathcal{D}$ la aplicación que transforma todo polinomio $f(\lambda)$ en su derivada $f'(\lambda)$. Demuéstrese que $\mathcal{D}^{n+1} = \mathcal{O}$. Hállese la matriz de $\mathcal{D}$ en el sistema de coordenadas $1, \lambda, \ldots, \lambda^n$.

2. Sea $\mathfrak{L}$ el espacio de dimensión infinita de todos los polinomios en λ. Indiquemos por $\mathcal{D}$ la operación de diferenciación y por $\mathcal{C}$ la operación de multiplicación de polinomios por λ. Demuéstrese que ambas operaciones son lineales y que están ligadas por las relaciones

$$\mathcal{C}^n\mathcal{D} - \mathcal{D}\mathcal{C}^n = n\mathcal{C}^{n-1} \quad (n = 1, 2, \ldots).$$

3. ¿Por qué no se puede considerar la aplicación $\mathcal{C}$, indicada en el problema anterior, en el espacio de polinomios de grado no mayor de n?

4. Las aplicaciones lineales de un espacio $\mathfrak{L}$ forman, respecto a las operaciones de adición y de multiplicación por número, un espacio lineal. ¿Cuál es la dimensión de este espacio, si la dimensión de $\mathfrak{L}$ es igual a n?

§ 10. Rango y defecto de una aplicación lineal

Hasta el momento hemos considerado aquellas propiedades de las aplicaciones lineales que se refieren principalmente a las reglas de las operaciones con las mismas. Estudiemos ahora algunas propiedades de carácter más bien geométrico.

10.1. Núcleo y dominio de valores. Sea $\mathfrak{M}$ un conjunto de vectores de un espacio lineal $\mathfrak{L}$ y sea $\mathcal{A}$ una aplicación lineal cualquiera del último. Todo vector a de $\mathfrak{M}$ se transforma por la aplicación lineal $\mathcal{A}$ en un nuevo vector $a\mathcal{A}$ que es la imagen del

vector a. En general, esta imagen no pertenecerá a $\mathfrak{M}$. El conjunto de las imágenes de todos los vectores de $\mathfrak{M}$ se llamará *imagen* de $\mathfrak{M}$ respecto de $\mathcal{A}$ y se indicará por $\mathfrak{M}\mathcal{A}$. Convendremos en llamar *imagen recíproca* del conjunto $\mathfrak{M}$ el conjunto de todos los vectores de $\mathfrak{L}$ cuyas imágenes pertenezcan a $\mathfrak{M}$.

TEOREMA 1. *Las imágenes y las imágenes recíprocas de los subespacios lineales de un espacio $\mathfrak{L}$ respecto a una aplicación lineal cualquiera $\mathcal{A}$ son también subespacios lineales.*

En efecto, sea $\mathfrak{N}$ un subespacio lineal de $\mathfrak{L}$. Mostremos que $\mathfrak{N}\mathcal{A}$ es también un subespacio lineal. Tomemos en $\mathfrak{N}\mathcal{A}$ unos vectores a y b cualesquiera. Estos vectores son las imágenes de unos vectores x e y de $\mathfrak{N}$, es decir, $a = x\mathcal{A}$ y $b = y\mathcal{A}$. Puesto que $\mathfrak{N}$ es un subespacio lineal, el vector $\alpha x + \beta y$ pertenece a $\mathfrak{N}$ cualesquiera que sean α y β. Por ello, el vector $(\alpha x + \beta y)\mathcal{A}$ también pertenece a $\mathfrak{N}\mathcal{A}$. Pero tenemos

$$(\alpha x + \beta y)\mathcal{A} = \alpha \cdot x\mathcal{A} + \beta \cdot y\mathcal{A} = \alpha a + \beta b,$$

es decir, $\mathfrak{N}\mathcal{A}$ contiene el vector $\alpha a + \beta b$ y, por consiguiente, es un subespacio lineal. De forma análoga se demuestra también que la imagen recíproca de un subespacio lineal $\mathfrak{N}$ es un subespacio lineal.

Se llama *núcleo* de una aplicación lineal $\mathcal{A}$ el conjunto de todos los vectores de $\mathfrak{L}$ que se transforman por la aplicación $\mathcal{A}$ en el vector nulo y se llama *dominio de valores* de $\mathcal{A}$ el conjunto de las imágenes de todos los vectores de $\mathfrak{L}$. La dimensión del dominio de valores se llama *rango* de la aplicación y la dimensión del núcleo se llama *defecto* de la aplicación.

TEOREMA 2. *La suma del rango y del defecto de una aplicación lineal $\mathcal{A}$ es igual a la dimensión del espacio $\mathfrak{L}$.*

Indiquemos por $\mathfrak{N}$ el núcleo de la aplicación $\mathcal{A}$ y supongamos que d es la dimensión del núcleo $\mathfrak{N}$. Indiquemos por r la dimensión del dominio de valores $\mathfrak{L}\mathcal{A}$. Por definición, d y r son, respectivamente, el defecto y el rango de la aplicación $\mathcal{A}$. Tomemos en el dominio de valores $\mathfrak{L}\mathcal{A}$ una base $a_1, a_2, \ldots, a_r$ y sean $b_1, b_2, \ldots, b_r$ unos elementos del espacio $\mathfrak{L}$ que se transforman por la aplicación $\mathcal{A}$ en $a_1, a_2, \ldots, a_r$, respectivamente. Los vectores $b_1, b_2, \ldots, b_r$ son linealmente independientes, ya que de la relación

$$\alpha_1 b_1 + \alpha_2 b_2 + \ldots + \alpha_r b_r = o$$

se deduce que

$$(\alpha_1 b_1 + \alpha_2 b_2 + \ldots + \alpha_r b_r)\mathcal{A} = \alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_r a_r = o,$$

y, puesto que $a_1, a_2, \ldots, a_r$ son linealmente independientes, se tiene

$$\alpha_1 = \alpha_2 = \ldots = \alpha_r = 0.$$

Consideremos el subespacio $\mathfrak{M}$ tendido sobre los vectores $b_1, b_2, \ldots, b_r$. El sistema $b_1, \ldots, b_r$ es *una base* de $\mathfrak{M}$ y por ello la

dimensión del subespacio $\mathfrak{M}$ es igual a r, es decir, al rango de la aplicación $\mathcal{A}$. Demostremos que el espacio $\mathfrak{L}$ es la suma directa de los subespacios $\mathfrak{M}$ y $\mathfrak{N}$. Para ello es suficiente demostrar, según el p. 6.2, que $\mathfrak{M} \cap \mathfrak{N} = \{o\}$ y que $\mathfrak{L} = \mathfrak{M} + \mathfrak{N}$. Demostremos lo primero. Todo vector de $\mathfrak{M}$ es de la forma

$$b = \alpha_1 b_1 + \alpha_2 b_2 + \ldots + \alpha_r b_r.$$

Si b pertenece a $\mathfrak{N}$, se tiene $b\mathcal{A} = o$, es decir,

$$(\alpha_1 b_1 + \ldots + \alpha_r b_r)\,\mathcal{A} = \alpha_1 a_1 + \ldots + \alpha_r a_r = o.$$

Pero los vectores $a_1, \ldots, a_r$ son linealmente independientes y por esto $\alpha_1 = \ldots = \alpha_r = 0$, de modo que $b = o$ que es lo que se quería demostrar. Resta probar que $\mathfrak{L} = \mathfrak{M} + \mathfrak{N}$. Tomemos un vector cualquiera a de $\mathfrak{L}$. Su imagen $a\mathcal{A}$ pertenece a $\mathfrak{L}\mathcal{A}$ y, por consiguiente, se expresa linealmente en términos de $a_1, \ldots, a_r$:

$$a\mathcal{A} = \alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_r a_r.$$

Sea

$$b = \alpha_1 b_1 + \alpha_2 b_2 + \ldots + \alpha_r b_r \quad \text{y} \quad a - b = c.$$

Puesto que

$$b\mathcal{A} = \alpha_1 \cdot b_1 \mathcal{A} + \ldots + \alpha_r \cdot b_r \mathcal{A} = \alpha_1 a_1 + \ldots + \alpha_r a_r = a\mathcal{A},$$

se tiene

$$c\mathcal{A} = (a - b)\,\mathcal{A} = a\mathcal{A} - b\mathcal{A} = o.$$

Por consiguiente, c pertenece a $\mathfrak{N}$. Es decir, tenemos

$$a = b + c \quad (b \in \mathfrak{M}, \quad c \in \mathfrak{N});$$

pero esto significa precisamente que

$$\mathfrak{L} = \mathfrak{M} + \mathfrak{N}.$$

Como que esta suma es directa, la dimensión del espacio $\mathfrak{L}$ es igual a la suma de las dimensiones de los subespacios $\mathfrak{M}$ y $\mathfrak{N}$, es decir, es igual a la suma del rango y del defecto de la aplicación $\mathcal{A}$. Hemos demostrado el teorema.

Consideremos un ejemplo. Sea $\mathfrak{N}$ el espacio corriente de los segmentos orientados que parten de un punto O. Tomemos un plano cualquiera $\mathfrak{P}$ que pase por el punto O e indiquemos por $\mathcal{P}$ la operación de proyección ortogonal sobre $\mathfrak{P}$. La aplicación $\mathcal{P}$ transforma todo el espacio $\mathfrak{N}$ en el plano $\mathfrak{P}$. Es decir, $\mathfrak{P}$ es el dominio de valores de la aplicación $\mathcal{P}$ y el rango de $\mathcal{P}$ es igual a 2. El núcleo de la aplicación $\mathcal{P}$ está compuesto por los vectores que pertenecen a la recta que pasa por el punto O y que es perpendicular al plano $\mathfrak{P}$, ya que solamente estos vectores son transformados por la aplicación $\mathcal{P}$ en el nulo. Por consiguiente el defecto de $\mathcal{P}$ es igual

a 1. La suma del rango y del defecto de la aplicación $\mathcal{P}$ es igual a 3 tal y como debe ser según el teorema 2.

10.2. Aplicaciones singulares y regulares. Más arriba (en el p. 9.1) hemos señalado que no toda aplicación es invertible. En lo sucesivo las aplicaciones lineales invertibles serán llamadas *regulares*, mientras que las aplicaciones no invertibles se llamarán *singulares*. En el p. 9.1 hemos encontrado en forma matricial las condiciones que garantizan que una aplicación lineal sea regular. Queremos dar ahora a estas condiciones un carácter geométrico.

TEOREMA 3. *Para que una aplicación lineal $\mathcal{A}$ de un espacio $\mathfrak{L}$ sea regular es necesario y suficiente que el núcleo de esta aplicación sea nulo, es decir, que el defecto de $\mathcal{A}$ sea igual a cero.*

DEMOSTRACION. Si la aplicación dada $\mathcal{A}$ es regular, todo vector debe tener sólo una imagen recíproca; en particular, sólo una imagen recíproca debe tener el vector nulo o. Puesto que o siempre es una imagen recíproca del vector nulo y puesto que en este caso el núcleo está compuesto sólo de un vector, éste será precisamente el vector o.

Viceversa, sea el defecto de $\mathcal{A}$ igual a cero. En virtud del teorema 2, de aquí se deduce que el rango de $\mathcal{A}$ es igual a la dimensión de $\mathfrak{L}$, es decir, que la dimensión del dominio de valores $\mathfrak{L}\mathcal{A}$ es igual a la dimensión de $\mathfrak{L}$. Por consiguiente, $\mathfrak{L}\mathcal{A} = \mathfrak{L}$; vemos, pues, que todo vector de $\mathfrak{L}$ es la imagen de un vector de $\mathfrak{L}$. Si demostramos que la aplicación $\mathcal{A}$ transforma diferentes vectores a y b en diferentes vectores, esto significará precisamente que la aplicación $\mathcal{A}$ es invertible (p. 7.3). Pero de $a\mathcal{A} = b\mathcal{A}$ se deduce que $(a-b)\mathcal{A} = o$. Puesto que, por hipótesis, el núcleo de $\mathcal{A}$ es nulo, tenemos $a-b=o$, es decir, $a=b$ que es lo que se necesitaba. Hemos demostrado el teorema.

La igualdad a cero del defecto de $\mathcal{A}$ equivale a la coincidencia del rango de $\mathcal{A}$ y de la dimensión del espacio $\mathfrak{L}$. Por ello, el teorema 3 puede ser enunciado también en la forma siguiente:

TEOREMA 4. *Para que una aplicación lineal $\mathcal{A}$ de un espacio $\mathfrak{L}$ sea regular es necesario y suficiente que el dominio de valores de $\mathcal{A}$ coincida con $\mathfrak{L}$, es decir, que el rango de $\mathcal{A}$ sea igual a la dimensión de $\mathfrak{L}$.*

Una aplicación biyectiva de un espacio lineal sobre otro se denomina *isomorfismo*, si transforma una suma de vectores del primer espacio en la suma de los vectores correspondientes del segundo espacio y si transforma, además, el producto de un número por un vector del primer espacio en el producto del mismo número por el vector correspondiente del segundo espacio (p. 4.3). Si ambos espacios lineales coinciden, obtenemos una aplicación isomorfa de un espacio lineal sobre sí mismo. Toda aplicación de este tipo se llama

automorfismo de un espacio lineal. La definición de automorfismo coincide, obviamente, con la definición de una aplicación lineal regular. Por consiguiente, las aplicaciones lineales regulares de un espacio $\mathfrak{L}$ pueden ser consideradas como automorfismos de este espacio. De la definición misma de automorfismo resulta que los automorfismos de un espacio $\mathfrak{L}$ son aquellas superposiciones del espacio $\mathfrak{L}$ sobre sí mismo que conservan todas sus propiedades geométricas, es decir, las propiedades que se enuncian en términos de las operaciones de adición y de multiplicación por número.

Consideremos dos aplicaciones lineales arbitrarias $\mathcal{A}$ y $\mathcal{B}$ de un espacio $\mathfrak{L}$. Convendremos en llamar estas aplicaciones *isomorfas* o *semejantes*, si existe un automorfismo $\mathcal{C}$ del espacio $\mathfrak{L}$ que transforma una aplicación en la otra.

Sea u un vector de $\mathfrak{L}$ y sea $v = u\mathcal{A}$. El automorfismo $\mathcal{C}$ transforma u en un vector x y v en un vector y. Se dice que $\mathcal{C}$ transforma la aplicación $\mathcal{A}$ en $\mathcal{B}$, si $x\mathcal{B} = y$ (véase la fig. 3). Puesto que $x = u\mathcal{C}$ e $y = v\mathcal{C}$, de la igualdad $x\mathcal{B} = y$ resulta

$$u\mathcal{C}\mathcal{B} = v\mathcal{C} \quad \text{y} \quad u\mathcal{C}\mathcal{B}\mathcal{C}^{-1} = v = u\mathcal{A}.$$

De aquí

$$\mathcal{C}\mathcal{B}\mathcal{C}^{-1} = \mathcal{A} \quad \text{y} \quad \mathcal{B} = \mathcal{C}^{-1}\mathcal{A}\mathcal{C}. \tag{1}$$

Por consiguiente, *una aplicación $\mathcal{B}$ es isomorfa a una aplicación $\mathcal{A}$ si, y sólo si, se obtiene transformando $\mathcal{A}$ por un automorfismo del espacio $\mathfrak{L}$, es decir, por una aplicación regular de este espacio.*

Tomemos en $\mathfrak{L}$ un sistema de coordenadas y sean A, B y C las matrices respectivas de las aplicaciones $\mathcal{A}$, $\mathcal{B}$ y $\mathcal{C}$. Entonces la igualdad (1) equivale a la relación matricial

$$B = C^{-1}AC$$

y llegamos a la siguiente conclusión: *para que dos aplicaciones lineales de un espacio $\mathfrak{L}$ sean isomorfas es necesario y suficiente que sus matrices sean semejantes.*

Debemos considerar las aplicaciones isomorfas como aplicaciones que tienen las mismas propiedades geométricas. De aquí la importancia de saber clasificar, salvo un isomorfismo, todas las aplicaciones lineales. Algebraicamente este problema equivale a la clasificación, salvo semejanza, de todas las matrices cuadradas de orden n. El problema de clasificación de todos los espacios lineales sobre un cuerpo dado se resuelve sin dificultad (véase el p. 4.3); en cambio, el problema de la clasificación de las aplicaciones lineales exige para su resolución un estudio más detallado de las propiedades de las aplicaciones lineales. Este problema quedará totalmente resuelto sólo en el capítulo siguiente.

Señalemos, para concluir, una propiedad más de las aplicaciones lineales isomorfas: *para que las aplicaciones lineales $\mathcal{A}$ y $\mathcal{B}$ sean isomorfas es necesario y suficiente que existan unos sistemas de coor-*

denadas en los que estas aplicaciones se representen por una misma matriz.

En efecto, si $\mathcal{A}$ y $\mathcal{B}$ son isomorfas, sus matrices están ligadas por la relación $A = CBC^{-1}$. Pero esta relación muestra que, si pasamos del sistema de coordenadas dado a otro de manera que la matriz del cambio sea C, la matriz de la aplicación $\mathcal{B}$ en el nuevo sistema de coordenadas será A, es decir, coincidirá con la matriz de la aplicación $\mathcal{A}$ calculada en el sistema antiguo de coordenadas. Viceversa, si la matriz A de la aplicación $\mathcal{A}$ en el sistema de coordenadas $a_1, \ldots, a_n$ coincide con la matriz de la aplicación $\mathcal{B}$ calculada en el sistema de coordenadas $a'_1, \ldots, a'_n$ y si C es la matriz del cambio del primer sistema de coordenadas por el segundo, tendremos que las matrices de las aplicaciones $\mathcal{A}$ y $\mathcal{B}$ en el segundo sistema de coordenadas serán iguales, respectivamente, a CAC^{-1} y a A, es decir, serán semejantes, que es lo que se quería demostrar.

10.3. Rango de la matriz de una aplicación. En el p. 10.1 hemos introducido el concepto del rango de una aplicación lineal. Por otro lado, conocemos del p. 5.2 el concepto del rango de una matriz que es el número máximo de sus filas linealmente independientes. Surge el problema sobre la relación que existe entre estos conceptos. La respuesta viene dada por el teorema siguiente:

TEOREMA 5. *El rango de una aplicación lineal cualquiera $\mathcal{A}$ de un espacio $\mathfrak{L}$ coincide con el rango de la matriz de esta aplicación.*

Sea $a_1, \ldots, a_n$ un sistema de coordenadas de $\mathfrak{L}$. El dominio de valores de la aplicación $\mathcal{A}$ está compuesto por las imágenes de los vectores del espacio $\mathfrak{L}$, es decir, por vectores de tipo

$$(\alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_n a_n)\,\mathcal{A} = \alpha_1 (a_1 \mathcal{A}) + \alpha_2 (a_2 \mathcal{A}) + \ldots + \alpha_n (a_n \mathcal{A}).$$

De aquí se desprende que el dominio de valores de la aplicación $\mathcal{A}$ es el subespacio tendido sobre los vectores $a_1\mathcal{A}, \ldots, a_n\mathcal{A}$. El rango de $\mathcal{A}$ es igual a la dimensión del dominio de valores y, por consiguiente, al número máximo de vectores linealmente independientes entre los vectores $a_1\mathcal{A}, \ldots, a_n\mathcal{A}$ (p. 6.1). Sea

$$\begin{aligned}
a_1\mathcal{A} &= \alpha_{11}a_1 + \alpha_{12}a_2 + \ldots + \alpha_{1n}a_n,\\
a_2\mathcal{A} &= \alpha_{21}a_1 + \alpha_{22}a_2 + \ldots + \alpha_{2n}a_n,\\
&\cdots\cdots\cdots\cdots\cdots\\
a_n\mathcal{A} &= \alpha_{n1}a_1 + \alpha_{n2}a_2 + \ldots + \alpha_{nn}a_n.
\end{aligned}$$

La matriz

$$A = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \ldots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \ldots & \alpha_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ \alpha_{n1} & \alpha_{n2} & \ldots & \alpha_{nn} \end{bmatrix}$$

es la matriz de la aplicación $\mathcal{A}$. Las filas de esta matriz son las

filas de coordenadas de los vectores $a_1\mathcal{A}, \ldots, a_n\mathcal{A}$. Luego, el número máximo de vectores linealmente independientes entre los vectores $a_1\mathcal{A}, \ldots, a_n\mathcal{A}$ es igual al número máximo de filas linealmente independientes de la matriz A, es decir, es igual a su rango que es lo que se quería demostrar.

Se llama *defecto* de una matriz cuadrada la diferencia entre su orden y su rango. De los teoremas 2 y 5 se desprende directamente que *el defecto de una aplicación lineal es igual al defecto de su matriz.*

Consideremos a título de ejemplo un sistema de n ecuaciones lineales homogéneas

$$\left.\begin{aligned} \xi_1\alpha_{11}+\xi_2\alpha_{21}+\ldots+\xi_n\alpha_{n1}&=0,\\ \xi_1\alpha_{12}+\xi_2\alpha_{22}+\ldots+\xi_n\alpha_{n2}&=0,\\ \cdots\cdots\cdots\cdots&\\ \xi_1\alpha_{1n}+\xi_2\alpha_{2n}+\ldots+\xi_n\alpha_{nn}&=0, \end{aligned}\right\} \tag{2}$$

con n incógnitas $\xi_1, \xi_2, \ldots, \xi_n$. Queremos estudiar las soluciones de este sistema. Geométricamente este problema puede ser interpretado de la forma siguiente. Tomemos un espacio lineal cualquiera $\mathfrak{L}$ de n dimensiones y escojamos en él un sistema de coordenadas $a_1, \ldots, a_n$. Convengamos en interpretar las magnitudes incógnitas $\xi_1, \ldots, \xi_n$ como las coordenadas de un vector x de $\mathfrak{L}$. Consideremos la aplicación lineal $\mathcal{A}$ de este espacio cuya matriz A está compuesta por los elementos α_{ij} $(i, j = 1, 2, \ldots, n)$. El sistema de ecuaciones (2) puede ser representado en la forma

$$[x]\,A = O$$

o en la forma vectorial

$$x\mathcal{A} = o.$$

De aquí se ve que las soluciones del sistema (2) son las filas de coordenada de los vectores pertenecientes al núcleo de la aplicación $\mathcal{A}$. Puesto que la dimensión del núcleo coincide con el defecto de la aplicación y el defecto de la aplicación es igual al defecto de su matriz, llegamos al conocido teorema de la teoría de los determinantes: *el número máximo de soluciones linealmente independientes de un sistema homogéneo de n ecuaciones lineales con n incógnitas es igual al defecto de la matriz de este sistema.*

Ejemplos y problemas

1. En el espacio de filas de longitud 4 se ha tomado el sistema de coordenadas

$$e_1 = [1, 0, 0, 0],\quad e_2 = [0, 1, 0, 0],\quad e_3 = [0, 0, 1, 0] \quad \text{y} \quad e_4 = [0, 0, 0, 1].$$

Hállense el núcleo y el dominio de valores de las aplicaciones lineales definidas por las matrices

$$\begin{bmatrix} 2 \\ 0 \\ 3 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 2 & 1 & 0 & 0 \\ 2 & 1 & 0 & \\ & 2 & 0 & \\ & & 0 & \end{bmatrix}, \quad \begin{bmatrix} 3 & 1 & & \\ 2 & 1 & & \\ & & 0 & 0 \\ & & 0 & 0 \end{bmatrix} \text{ y } \begin{bmatrix} 1 & 3 & 5 & 1 \\ 2 & 1 & 3 & 1 \\ 4 & 7 & 13 & 3 \\ 3 & -1 & 1 & 1 \end{bmatrix}.$$

2. Sea $\mathcal{A}$ una aplicación lineal de un espacio $\mathfrak{L}$ y sea $\mathfrak{M}$ un subespacio lineal de $\mathfrak{L}$. Demuéstrese que la dimensión de la imagen del subespacio $\mathfrak{M}$ satisface las desigualdades

$$\dim.\,\mathfrak{M} - \text{def.}\,\mathcal{A} \leqslant \dim.\,\mathfrak{M}\mathcal{A} \leqslant \dim.\,\mathfrak{M}.$$

3. Si $\mathcal{A}_1$ y $\mathcal{A}_2$ son unas aplicaciones lineales arbitraria y regular, respectivamente, se tiene

$$\text{rango } (\mathcal{A}_1\mathcal{A}_2) = \text{rango } (\mathcal{A}_2\mathcal{A}_1) = \text{rango } \mathcal{A}_1.$$

4. Sean $\mathcal{A}_1$ y $\mathcal{A}_2$ unas aplicaciones lineales cualesquiera de un espacio lineal $\mathfrak{L}$. Entonces se tiene

$$\text{rango } (\mathcal{A}_1 + \mathcal{A}_2) \leqslant \text{rango } \mathcal{A}_1 + \text{rango } \mathcal{A}_2,$$
$$\text{def. } (\mathcal{A}_1\mathcal{A}_2) \leqslant \text{def. } \mathcal{A}_1 + \text{def. } \mathcal{A}_2,$$
$$\text{rango } (\mathcal{A}_1\mathcal{A}_2) \leqslant \text{rango } \mathcal{A}_1 \text{ y rango } (\mathcal{A}_1\mathcal{A}_2) \leqslant \text{rango } \mathcal{A}_2.$$

5. Toda aplicación lineal de rango m puede ser representada en forma de una suma de aplicaciones de rango 1.

6. Para que una matriz A de orden n tenga el rango no mayor que 1 es necesario y suficiente que A pueda ser representada en la forma

$$A = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} (\beta_1\beta_2 \ldots \beta_n) = \| \alpha_i\beta_j \|,$$

donde α_i y β_i son números determinados.

§ 11. Subespacios invariantes

11.1. Aplicación inducida. Se dice que un subespacio $\mathfrak{A}$ de un espacio lineal $\mathfrak{L}$ *es invariante* respecto a una aplicación $\mathcal{A}$ si todo vector de $\mathfrak{A}$ se transforma por la aplicación $\mathcal{A}$ de nuevo en un vector de $\mathfrak{A}$, es decir, si

$$\mathfrak{A}\mathcal{A} \subset \mathfrak{A}.$$

De esta definición se desprende directamente que los subespacios impropios (el subespacio nulo y el propio espacio $\mathfrak{L}$) son invariantes respecto a cualquier aplicación lineal. También directamente se deduce la proposición de que *toda suma y toda intersección de subespacios invariantes es de nuevo un subespacio invariante.*

Observemos además que *siendo un subespacio $\mathfrak{A}$ invariante respecto a una aplicación $\mathcal{A}$, $\mathfrak{A}$ también será invariante respecto a la aplicación $f(\mathcal{A})$, donde $f(\mathcal{A}) = \alpha_0\mathcal{E} + \alpha_1\mathcal{A} + \ldots + \alpha_m\mathcal{A}^m$ es un polinomio arbitrario en $\mathcal{A}$.*

En efecto, si a es un vector de $\mathfrak{A}$, el vector $a\mathcal{A}$ está contenido, por hipótesis, en $\mathfrak{A}$. De aquí se deduce que $(a\mathcal{A})\mathcal{A} = a\mathcal{A}^2$ está contenido en $\mathfrak{A}$, etc. Vemos, por consiguiente, que el vector $a\mathcal{A}^k$ pertenece a $\mathfrak{A}$ cualquiera que sea $k \geqslant 0$. Pero en este caso $\mathfrak{A}$ contiene también todas las combinaciones lineales de estos vectores; en particular, $\mathfrak{A}$ contiene el vector

$$af(\mathcal{A}) = \alpha_0 a + \alpha_1 \cdot a\mathcal{A} + \ldots + \alpha_m \cdot a\mathcal{A}^m,$$

que es lo que se quería demostrar.

Los métodos de determinación de los subespacios invariantes serán considerados en el p. 11.4 y en el § 12; ahora, queremos

sólo explicar cómo se pueden aprovechar los subespacios invariantes para simplificar la matriz de una aplicación.

Sea $\mathfrak{A}$ un subespacio invariante no trivial de una aplicación lineal $\mathcal{A}$. Tomemos en $\mathfrak{A}$ una base $a_1, a_2, \ldots, a_m$ complementándola con vectores linealmente independientes $a_{m+1}, \ldots, a_n$ hasta obtener una base de todo el espacio $\mathfrak{L}$. Para hallar la matriz de la aplicación $\mathcal{A}$ en el sistema de coordenadas $a_1, \ldots, a_m, a_{m+1}, \ldots, a_n$ es necesario expresar linealmente los vectores $a_1\mathcal{A}, \ldots, a_n\mathcal{A}$ en términos de los vectores coordenados $a_1, \ldots, a_n$. Pero el subespacio $\mathfrak{A}$ es invariante y, por ello, los vectores $a_1\mathcal{A}, \ldots, a_m\mathcal{A}$ pertenecen de nuevo a $\mathfrak{A}$ y se expresan linealmente en términos de $a_1, \ldots, a_m$. Por consiguiente, tenemos

$$\left.\begin{aligned}
a_1\mathcal{A} &= \alpha_{11}a_1 + \ldots + \alpha_{1m}a_m,\\
&\cdots\cdots\cdots\cdots\cdots\\
a_m\mathcal{A} &= \alpha_{m1}a_1 + \ldots + \alpha_{mm}a_m,\\
a_{m+1}\mathcal{A} &= \alpha_{m+1,1}a_1 + \ldots + \alpha_{m+1,m}a_m + \ldots + \alpha_{m+1,n}a_n,\\
&\cdots\cdots\cdots\cdots\cdots\\
a_n\mathcal{A} &= \alpha_{n1}a_1 + \ldots + \alpha_{nm}a_m + \ldots + \alpha_{nn}a_n,
\end{aligned}\right\} \qquad (1)$$

es decir, la matriz de la aplicación $\mathcal{A}$ es igual a

$$A = \begin{bmatrix}
\alpha_{11} & \ldots & \alpha_{m1} & 0 & \ldots & 0\\
\cdot & & \cdot & \cdot & & \cdot\\
\cdot & & \cdot & \cdot & & \cdot\\
\cdot & & \cdot & \cdot & & \cdot\\
\alpha_{m1} & \ldots & \alpha_{mm} & 0 & \ldots & 0\\
\alpha_{m+1,1} & \ldots & \alpha_{m+1,m} & \alpha_{m+1,m+1} & \ldots & \alpha_{m+1,n}\\
\cdot & & \cdot & \cdot & & \cdot\\
\cdot & & \cdot & \cdot & & \cdot\\
\cdot & & \cdot & \cdot & & \cdot\\
\alpha_{n1} & \ldots & \alpha_{nm} & \alpha_{n,m+1} & \ldots & \alpha_{nn}
\end{bmatrix} = \begin{bmatrix} A_1 & O\\ B & A_2 \end{bmatrix}. \qquad (2)$$

Resumiendo, *si una aplicación lineal posee un subespacio lineal invariante, su matriz en un sistema de coordenadas adecuado se descompone en cuatro células con la particularidad de que las células diagonales son cuadradas, mientras que la célula superior de la derecha puede resultar rectangular, pero formada íntegramente por ceros.* Las matrices de este tipo han sido llamadas en el p. 2.1 *semidescompuestas*. Recíprocamente, si en un sistema de coordenadas la matriz de una aplicación lineal $\mathcal{A}$ tiene la forma semidescompuesta (2), las igualdades (1) muestran que en este caso el subespacio $\mathfrak{A}$ tendido sobre los m primeros vectores coordenados será invariante respecto de $\mathcal{A}$.

Geométricamente la matriz A_1 puede ser interpretada de modo siguiente. Tenemos que la aplicación $\mathcal{A}$ transforma todo vector de $\mathfrak{A}$ de nuevo en un vector de $\mathfrak{A}$. Por esto $\mathcal{A}$ puede ser considerada también como una aplicación del espacio $\mathfrak{A}$. Indiquemos esta aplicación por $\mathcal{A}_1$ y convengamos en llamar $\mathcal{A}_1$ aplicación *inducida*. Las aplicaciones $\mathcal{A}$ y $\mathcal{A}_1$ actúan sobre los vectores del subespacio $\mathfrak{A}$ idénticamente: si a es un vector de $\mathfrak{A}$, se tiene $a\mathcal{A} = a\mathcal{A}_1$. La diferencia entre estas aplicaciones consiste en que tienen distintos campos de definición: si a es un vector del espacio principal $\mathfrak{L}$ que no pertenece a $\mathfrak{A}$, la operación $a\mathcal{A}$ tiene sentido, mientras que $a\mathcal{A}_1$ no lo tiene.

Las primeras m igualdades del sistema (1) muestran que A_1 *es la matriz de la aplicación inducida* $\mathcal{A}_1$ en el sistema de coordenadas $a_1, a_2, \ldots, a_m$.

11.2. Suma directa de subespacios invariantes. Hemos considerado el caso en que la aplicación lineal $\mathcal{A}$ tiene sólo un subespacio invariante. Supongamos ahora que $\mathcal{A}$ posee dos subespacios invariantes $\mathfrak{A}_1$ y $\mathfrak{A}_2$ y, es más, supongamos que el espacio $\mathfrak{L}$ es la suma directa de estos subespacios. Tomemos en $\mathfrak{A}_1$ y $\mathfrak{A}_2$ unos sistemas de coordenadas $a_1, \ldots, a_m$ y $a_{m+1}, \ldots, a_n$, respectivamente. Según el p. 6.2, los vectores $a_1, \ldots, a_m, a_{m+1}, \ldots, a_n$ forman un sistema de coordenadas de $\mathfrak{L}$. Veamos la forma que toma en este sistema la matriz de la aplicación $\mathcal{A}$. Por hipótesis, los vectores $a_1\mathcal{A}, \ldots, a_m\mathcal{A}$ pertenecen a $\mathfrak{A}_1$ y los vectores $a_{m+1}\mathcal{A}, \ldots, a_n\mathcal{A}$ pertenecen a $\mathfrak{A}_2$. Luego, tenemos

$$\left.\begin{aligned}
a_1\mathcal{A} &= \alpha_{11}a_1 + \ldots + \alpha_{1m}a_m,\\
&\cdots\cdots\cdots\cdots\\
a_m\mathcal{A} &= \alpha_{m1}a_1 + \ldots + \alpha_{mm}a_m,\\
a_{m+1}\mathcal{A} &= \alpha_{m+1,\,m+1}a_{m+1} + \ldots + \alpha_{m+1,n}a_n,\\
&\cdots\cdots\cdots\cdots\\
a_n\mathcal{A} &= \alpha_{n,\,m+1}a_{m+1} + \ldots + \alpha_{nn}a_n.
\end{aligned}\right\} \qquad (3)$$

Por consiguiente, la matriz de la aplicación $\mathcal{A}$ resulta ser igual a

$$A = \begin{bmatrix}
\alpha_{11} & \ldots & \alpha_{1m} & 0 & \ldots & 0\\
\vdots & & \vdots & \vdots & & \vdots\\
\alpha_{m1} & \ldots & \alpha_{mm} & 0 & \ldots & 0\\
0 & \ldots & 0 & \alpha_{m+1,\,m+1} & \ldots & \alpha_{m+1,\,n}\\
\vdots & & \vdots & \vdots & & \vdots\\
0 & \ldots & 0 & \alpha_{n,\,m+1} & \ldots & \alpha_{nn}
\end{bmatrix} = \begin{bmatrix} A_1 & O\\ O & A_2 \end{bmatrix}. \qquad (4)$$

es decir, resulta ser descompuesta. Sean $\mathcal{A}_1$ y $\mathcal{A}_2$ las aplicaciones inducidas en los subespacios $\mathfrak{A}_1$ y $\mathfrak{A}_2$ por la aplicación $\mathcal{A}$. De las igualdades (3) se deduce que A_1 y A_2 son las matrices de las aplicaciones $\mathcal{A}_1$ y $\mathcal{A}_2$ en los sistemas correspondientes de coordenadas.

Resumiendo, *si un espacio* $\mathfrak{L}$ *se descompone en una suma directa de subespacios invariantes respecto a una aplicación lineal* $\mathcal{A}$, *la matriz de la aplicación* $\mathcal{A}$ *calculada en un sistema adecuado de coordenadas toma la forma celular diagonal y sus células diagonales representan las matrices de las aplicaciones inducidas por la aplicación* $\mathcal{A}$ *en los subespacios invariantes.*

Hemos demostrado esta proposición sólo en el caso de una suma de dos subespacios invariantes. Sin embargo, todos los razonamientos se traspasan sin modificaciones al caso de un número arbitrario de sumandos.

Supongamos ahora lo contrario aceptando que en un sistema de coordenadas la matriz de la aplicación $\mathcal{A}$ toma la forma descompuesta (4). Entonces de las igualdades (3) se desprende que el espacio $\mathfrak{A}_1$ tendido sobre los m primeros vectores coordenados y el espacio $\mathfrak{A}_2$ tendido sobre los restantes vectores coordenados serán invariantes respecto de $\mathcal{A}$. Es obvio que la suma $\mathfrak{A}_1 + \mathfrak{A}_2$ es directa y coincide con $\mathfrak{L}$. Por consiguiente, la condición de que la matriz de la aplicación $\mathcal{A}$ se reduce a la forma celular diagonal, además de ser necesaria, es también suficiente para que $\mathfrak{L}$ sea la suma directa de los subespacios invariantes respecto de $\mathcal{A}$.

Examinemos el problema siguiente. Se tiene una descomposición de un espacio $\mathfrak{L}$ en la suma directa de unos subespacios lineales

$$\mathfrak{L} = \mathfrak{A}_1 \dot{+} \mathfrak{A}_2 \dot{+} \ldots \dot{+} \mathfrak{A}_s$$

y en cada subespacio $\mathfrak{A}_i$ se tiene una aplicación lineal $\mathcal{A}_i$. ¿Existe una aplicación lineal $\mathcal{A}$ del espacio $\mathfrak{L}$ respecto a la cual todos los subespacios $\mathfrak{A}_i$ son invariantes y que induce en todo $\mathfrak{A}_i$ la aplicación $\mathcal{A}_i$? ¿Será esta aplicación única? La respuesta es, obviamente, afirmativa. En efecto, tomemos en cada uno de los subespacios $\mathfrak{A}_i$ una base $a_{i1}, a_{i2}, \ldots, a_{im_i}$ e indiquemos por A_i la matriz de la aplicación $\mathcal{A}_i$ en este sistema de coordenadas. Consideremos la matriz celular diagonal

$$A = A_1 \dot{+} A_2 \dot{+} \ldots \dot{+} A_s.$$

El sistema formado por los vectores $a_{11}, \ldots, a_{1m_1}, \ldots, a_{s1}, \ldots, a_{sm_s}$ es una base del espacio $\mathfrak{L}$. En esta base a la matriz A le corresponde una aplicación lineal $\mathcal{A}$ del espacio $\mathfrak{L}$. En virtud de lo expuesto anteriormente, la aplicación $\mathcal{A}$ satisface todas las condiciones de nuestro problema. Esta aplicación es única, ya que su matriz en el sistema de coordenadas indicado se determina unívocamente por las condiciones del problema.

11.3. Polinomio característico de una aplicación. En este punto, así como en el siguiente, se supone que el cuerpo principal es un cuerpo conmutativo. Tomemos una aplicación lineal cualquiera $\mathcal{A}$ de un espacio lineal $\mathfrak{L}$ de n dimensiones. Escogiendo en $\mathfrak{L}$ un sistema de coordenadas determinado $a_1, \ldots, a_n$ podemos calcular la matriz A de la aplicación $\mathcal{A}$. El polinomio característico $\varphi(\lambda) = |\lambda E - A|$ de la matriz A se llama *polinomio característico de la aplicación* $\mathcal{A}$. Si tomamos otro sistema de coordenadas $a'_1, \ldots, a'_n$ e indicamos por T la matriz del cambio, la matriz de la aplicación $\mathcal{A}$ en el nuevo sistema de coordenadas será, de acuerdo con el p. 8.3, la matriz

$$A_1 = TAT^{-1},$$

es decir, la matriz semejante de A. Sin embargo, en el p. 3.2 hemos demostrado que las matrices semejantes tienen los mismos polinomios característicos. Por consiguiente, *el polinomio característico de una aplicación* $\mathcal{A}$ *no depende del sistema de coordenadas en el que se calcula.*

El grado del polinomio característico es igual al orden de la matriz A y el orden de la matriz A es igual a la dimensión del espacio $\mathfrak{L}$. Por esto, *el grado del polinomio característico de una aplicación* $\mathcal{A}$ *es igual a la dimensión del espacio en el que actúa esta aplicación.*

La suma de las raíces del polinomio característico es igual a la traza y el producto de las raíces es igual al determinante de la matriz A. Puesto que el polinomio característico de $\mathcal{A}$ y, por consiguiente, también sus raíces no dependen del sistema de coordenadas, tampoco la traza y el determinante de A dependerán del sistema de coordenadas. Por esta razón la traza y el determinante de la matriz de una aplicación $\mathcal{A}$ se llaman *traza y determinante de la aplicación* $\mathcal{A}$.

Si el espacio $\mathfrak{L}$ se descompone en una suma directa de los subespacios $\mathfrak{A}_1$ y $\mathfrak{A}_2$ invariantes respecto de $\mathcal{A}$, la matriz de la aplicación $\mathcal{A}$ toma, en un sistema de coordenadas adecuado, la forma celular diagonal

$$A = \begin{bmatrix} A_1 & O \\ O & A_2 \end{bmatrix}.$$

Según el p. 3.3. el polinomio característico de la matriz A es igual en este caso al producto de los polinomios característicos de las matrices A_1 y A_2. Pero A_1 y A_2 son las matrices de las aplicaciones lineales inducidas por la aplicación $\mathcal{A}$ en los subespacios invariantes $\mathfrak{A}_1$ y $\mathfrak{A}_2$. Por lo tanto, *si un espacio* $\mathfrak{L}$ *se descompone en la suma directa de subespacios invariantes respecto a una aplicación lineal* $\mathcal{A}$, *el polinomio característico de la aplicación* $\mathcal{A}$ *es igual al*

producto de los polinomios característicos de las aplicaciones inducidas por la aplicación $\mathcal{A}$ en los subespacios invariantes.

Según el teorema de Hamilton—Cayley (p. 3.2), toda matriz cuadrada A es raíz de su polinomio característico $\varphi(\lambda)$, es decir, $\varphi(A)=O$. Sea $\mathcal{A}$ una aplicación lineal de matriz A. La matriz de la aplicación $\varphi(\mathcal{A})$ es, de acuerdo con el p. 9.3, $\varphi(A)$. Como que esta matriz es nula, tenemos $\varphi(\mathcal{A})=\mathcal{O}$. Por consiguiente, *toda aplicación lineal es raíz de su polinomio característico.*

Se llama *polinomio mínimo de una aplicación lineal* $\mathcal{A}$ el polinomio de menor grado de coeficiente principal igual a 1 para el que la aplicación $\mathcal{A}$ es una raíz. Sea A la matriz de la aplicación $\mathcal{A}$ calculada en un sistema de coordenadas. Puesto que las relaciones $f(A)=O$ y $f(\mathcal{A})=\mathcal{O}$, donde $f(\lambda)$ es un polinomio arbitrario, son equivalentes, resulta que *el polinomio mínimo de una aplicación coincide con el polinomio mínimo de la matriz de esta aplicación.*

Si el espacio $\mathfrak{L}$ se descompone en suma directa de subespacios invariantes respecto a una aplicación $\mathcal{A}$, la matriz de la aplicación $\mathcal{A}$, en un sistema de coordenadas adecuado, se descompone. El polinomio mínimo de una matriz descompuesta es el mínimo común múltiplo de los polinomios mínimos de sus células diagonales (p. 3.3). Por esto, el polinomio mínimo de una aplicación $\mathcal{A}$ será igual al mínimo común múltiplo de los polinomios mínimos de las aplicaciones inducidas por la aplicación $\mathcal{A}$ en los subespacios invariantes.

11.4. Vectores propios y valores propios. Continuemos suponiendo que el cuerpo principal es un cuerpo conmutativo. Queremos ahora estudiar más detalladamente los subespacios invariantes de una dimensión. Introduzcamos primero la definición siguiente. Un número ζ se llama *valor propio* de una aplicación lineal $\mathcal{A}$, si existe en el espacio $\mathfrak{L}$ un vector no nulo a, tal que

$$a\mathcal{A}=\zeta a. \tag{5}$$

Todo vector que satisface esta relación se llama *vector propio* de la aplicación $\mathcal{A}$ correspondiente al valor propio ζ.

La búsqueda de los vectores propios y la búsqueda de los subespacios invariantes de una dimensión son problemas equivalentes. Efectivamente, sea a un vector no nulo propio de una aplicación $\mathcal{A}$ y sea ζ su valor propio correspondiente. Consideremos el subespacio de una dimensión $\mathfrak{A}$ tendido sobre el vector a, es decir, el conjunto de todos los vectores de tipo αa. La relación

$$(\alpha a)\,\mathcal{A}=\alpha\,(a\mathcal{A})=\zeta\alpha\cdot a \tag{6}$$

muestra que $\mathfrak{A}$ es invariante respecto de $\mathcal{A}$. Recíprocamente, sea $\mathfrak{A}$ un subespacio de dimensión uno invariante respecto de $\mathcal{A}$. Tomemos en $\mathfrak{A}$ un vector arbitrario no nulo a. Puesto que $\mathfrak{A}$ es de una dimensión, todos los vectores de $\mathfrak{A}$ son de la forma αa. Por hipó-

tesis, $a\mathcal{A}$ pertenece a $\mathfrak{A}$; luego,

$$a\mathcal{A} = \zeta a,$$

es decir, a es un vector propio de la aplicación $\mathcal{A}$ correspondiente al valor propio ζ. La igualdad (6) muestra que todos los demás vectores de $\mathfrak{A}$ también son vectores propios correspondientes al valor propio ζ.

Escojamos en el espacio $\mathfrak{L}$ un sistema de coordenadas $a_1, \ldots, a_n$ y sea $A = \|\alpha_{ij}\|$ la matriz de una aplicación lineal $\mathcal{A}$ en este sistema. Indiquemos por a algún vector propio no nulo de la aplicación $\mathcal{A}$. Sea $[\xi_1, \xi_2, \ldots, \xi_n]$ la fila de coordenadas del vector a y sea ζ el valor propio correspondiente. Pasando en la igualdad (5) a las coordenadas, obtenemos

$$[a]\, A = \zeta\, [a], \tag{7}$$

o

$$\left.\begin{aligned} \xi_1\alpha_{11} + \xi_2\alpha_{21} + \ldots + \xi_n\alpha_{n1} &= \zeta\xi_1, \\ \xi_1\alpha_{12} + \xi_2\alpha_{22} + \ldots + \xi_n\alpha_{n2} &= \zeta\xi_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots & \\ \xi_1\alpha_{1n} + \xi_2\alpha_{2n} + \ldots + \xi_n\alpha_{nn} &= \zeta\xi_n. \end{aligned}\right\} \tag{8}$$

Pasando todos los términos a un mismo miembro, obtenemos

$$\left.\begin{aligned} \xi_1(\zeta - \alpha_{11}) - \xi_2\alpha_{21} - \ldots - \xi_n\alpha_{n1} &= 0, \\ -\xi_1\alpha_{12} + \xi_2(\zeta - \alpha_{22}) - \ldots - \xi_n\alpha_{n2} &= 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots & \\ -\xi_1\alpha_{1n} - \xi_2\alpha_{2n} - \ldots + \xi_n(\zeta - \alpha_{nn}) &= 0. \end{aligned}\right\} \tag{9}$$

Este sistema puede ser considerado como un sistema de n ecuaciones lineales homogéneas con n incógnitas $\xi_1, \xi_2, \ldots, \xi_n$. Puesto que las coordenadas del vector *no nulo* propio a satisfacen el sistema (9), tenemos (véase el p. 10.3)

$$\begin{vmatrix} \zeta - \alpha_{11} & -\alpha_{12} & \ldots & -\alpha_{1n} \\ -\alpha_{21} & \zeta - \alpha_{22} & \ldots & -\alpha_{2n} \\ \ldots\ldots\ldots & \ldots\ldots\ldots & \ldots & \ldots\ldots\ldots \\ -\alpha_{n1} & -\alpha_{n2} & \ldots & \zeta - \alpha_{nn} \end{vmatrix} = |\zeta E - A| = 0, \tag{10}$$

donde E es la matriz unidad. Pero $|\lambda E - A|$ es el polinomio característico de la matriz A; por ello, de la igualdad (10) se desprende que *todos los valores propios de una aplicación lineal son raíces de su polinomio característico.* Recíprocamente, *si ζ es una raíz del polinomio característico de una aplicación $\mathcal{A}$ que pertenece al cuerpo conmutativo de coeficientes del espacio lineal, resulta que ζ es un valor propio de la aplicación $\mathcal{A}$.* En efecto, la igualdad (10) muestra que el rango de la matriz del sistema (9) será menor que n; por

consiguiente, este sistema tendrá al menos una solución no nula. Indicando esta solución por $[\xi_1, \xi_2, \ldots, \xi_n]$, obtenemos directamente de (8) y de (7) que el vector a de coordenadas $\xi_1, \xi_2, \ldots, \xi_n$ será el vector propio no nulo deseado [1].

Se llama *multiplicidad* de un valor propio ζ de una aplicación lineal $\mathcal{A}$ la multiplicidad que tiene ζ como raíz del polinomio característico de la aplicación $\mathcal{A}$.

Consideremos un ejemplo. Supongamos que en el espacio lineal real de tres dimensiones de base a_1, a_2 y a_3 actúa una aplicación lineal $\mathcal{A}$ de matriz

$$A = \begin{bmatrix} 3 & 3 & 2 \\ 1 & 1 & -2 \\ -3 & -1 & 0 \end{bmatrix}.$$

Se necesita hallar los valores propios y los vectores propios de la aplicación $\mathcal{A}$. Calculamos, ante todo, el polinomio característico de la aplicación $\mathcal{A}$:

$$\begin{bmatrix} \lambda-3 & -3 & -2 \\ -1 & \lambda-1 & 2 \\ 3 & 1 & \lambda \end{bmatrix} = (\lambda-4)(\lambda^2+4).$$

Sus raíces son iguales a $\lambda_1 = 4$, $\lambda_2 = 2i$ y $\lambda_3 = -2i$. Puesto que el cuerpo principal es real, los dos últimos valores deben ser omitidos, mientras que el primer valor $\lambda_1 = 4$ será el valor propio buscado. Para hallar los vectores propios formamos el sistema (9) que, en nuestro caso, se convierte en

$$\left.\begin{aligned} \xi_1 - \xi_2 + 3\xi_3 &= 0, \\ -3\xi_1 + 3\xi_2 + \xi_3 &= 0, \\ -2\xi_1 + 2\xi_2 + 4\xi_3 &= 0. \end{aligned}\right\}$$

Resolviendo este sistema obtenemos

$$\xi_2 = \xi_1 \quad \text{y} \quad \xi_3 = 0,$$

donde ξ_1 es arbitrario. Por consiguiente, el vector $\xi_1 a_1 + \xi_1 a_2$ será un vector propio de la aplicación $\mathcal{A}$ cualquiera que sea ξ_1.

---

[1] Existe una demostración más breve de la última proposición. La condición $a\mathcal{A} = \zeta a$ puede ser representada en la forma $a(\zeta\mathcal{E} - \mathcal{A}) = o$. Esto demuestra que los vectores propios de $\mathcal{A}$ son vectores que pertenecen al núcleo de la aplicación $\zeta\mathcal{E} - \mathcal{A}$. Pero, para que el núcleo contenga vectores no nulos es necesario y suficiente que la aplicación sea singular (p. 10.2), es decir, es necesario y suficiente que sea $|\zeta E - A| = 0$.

Ejemplos y problemas

1. En un espacio complejo $\mathfrak{L}$ de base a_1, a_2, a_3 y a_4 está definida una aplicación $\mathcal{A}$ de matriz

$$A=\begin{bmatrix} 1 & 0 & 2 & -1 \\ 0 & 1 & 4 & -2 \\ 2 & -1 & 0 & 1 \\ 2 & -1 & -1 & 2 \end{bmatrix}.$$

Hállense los vectores propios y los valores propios de la aplicación $\mathcal{A}$. Demuéstrese que el subespacio tendido sobre los vectores $2a_1-a_2$, y $-a_3+a_4$ es invariante respecto de $\mathcal{A}$.

2. Supongamos que un espacio $\mathfrak{L}$ tiene una base formada por los vectores propios de una aplicación $\mathcal{A}$. ¿Cuál será la matriz de la aplicación $\mathcal{A}$ en esta base?

3. Si en un espacio $\mathfrak{L}$ con un sistema de coordenadas $a_1, a_2, \ldots, a_n$ la matriz de una aplicación lineal $\mathcal{A}$ es de forma celular semidescompuesta

$$A=\begin{bmatrix} A_1 & B \\ O & A_2 \end{bmatrix},$$

donde A_2 es una matriz cuadrada de orden m, el subespacio tendido sobre los m últimos vectores coordenados $a_{n-m+1}, \ldots, a_n$ será invariante respecto de $\mathcal{A}$.

4. Si la aplicación $\mathcal{A}$ es regular, todo subespacio invariante respecto de $\mathcal{A}$ también será invariante respecto de $\mathcal{A}^{-1}$.

5. Si un subespacio $\mathfrak{A}$ es invariante respecto a una aplicación lineal $\mathcal{A}$, la imagen y la imagen recíproca del subespacio $\mathfrak{A}$ también serán invariantes respecto de $\mathcal{A}$.

6. En un espacio lineal complejo toda aplicación lineal tiene al menos un vector propio no nulo.

7. Supongamos que en un sistema de coordenadas $a_1, \ldots, a_n$ la matriz de una aplicación $\mathcal{A}$ es de forma diagonal con diferentes elementos diagonales. Hállense todos los subespacios invariantes de la aplicación $\mathcal{A}$ y demuéstrese que el número de los mismos es igual a 2^n.

8. Si una aplicación lineal $\mathcal{A}$ de un espacio $\mathfrak{L}$ de n dimensiones tiene n valores propios diferentes, la matriz de la aplicación $\mathcal{A}$ se reduce, en un sistema de coordenadas adecuado, a la forma diagonal.

§ 12. Aplicaciones de matrices de forma normal

En este parágrafo serán examinadas las propiedades de las aplicaciones lineales, cuyas matrices tienen, en un sistema de coordenadas fijo, la así llamada forma normal de Jordan. Por consiguiente, supondremos de antemano que las matrices de las aplicaciones consideradas pueden ser reducidas a esta forma. Más adelante, en el p. 15.4, veremos que esta reducción es siempre posible en el cuerpo de los números complejos.

En todo este parágrafo se supone que el cuerpo principal es un cuerpo conmutativo.

12.1. Forma diagonal. El teorema que sigue ofrece la característica más simple de las aplicaciones, cuyas matrices pueden ser reducidas a la forma diagonal.

TEOREMA 1 *Si una aplicación lineal de un espacio de n dimensiones tiene n vectores propios linealmente independientes, entonces tomándolos como vectores coordenados reduciremos la matriz de la aplicación a la forma diagonal. Recíprocamente, si la matriz de una aplicación en un sistema de coordenadas es de forma diagonal, los vectores de la base son vectores propios de la aplicación.*

La demostración es evidente. Un problema más sutil consiste en averiguar, a partir de la matriz de una aplicación calculada en un sistema de coordenadas eventual, si la aplicación posee vectores propios que constituyan una base del espacio. Este problema quedará resuelto en el p. 15.4, mientras que ahora estudiaremos sólo un caso particular del mismo.

TEOREMA 2 *Los vectores propios correspondientes a diferentes valores propios de una aplicación lineal son linealmente independientes.*

En efecto, sean $\rho_1, \rho_2, \ldots, \rho_m$ distintos valores propios y sean $a_1, a_2, \ldots, a_m$ los vectores propios que les corresponden de una aplicación lineal $\mathcal{A}$. Por inducción, podemos aceptar que $a_1, \ldots, a_{m-1}$ son linealmente independientes. Supongamos que $a_1, \ldots, a_m$ están ligados por una relación

$$\alpha_1 a_1 + \ldots + \alpha_{m-1} a_{m-1} + \alpha_m a_m = 0.$$

Aplicando a ambos miembros de esta igualdad la aplicación $\mathcal{A}$, obtenemos

$$\alpha_1 \rho_1 a_1 + \ldots + \alpha_{m-1} \rho_{m-1} a_{m-1} + \alpha_m \rho_m a_m = 0.$$

Eliminando a_m, tendremos

$$\alpha_1 (\rho_m - \rho_1) a_1 + \ldots + \alpha_{m-1} (\rho_m - \rho_{m-1}) a_{m-1} = 0.$$

Debido a la independencia lineal de $a_1, \ldots, a_{m-1}$ de aquí resulta $\alpha_1 = \ldots = \alpha_{m-1} = 0$ y, por consiguiente, $\alpha_m = 0$ que es lo que se quería demostrar.

Comparando ambos teoremas demostrados, obtenemos el corolario siguiente: *si el polinomio característico de una aplicación lineal de un espacio de n dimensiones tiene n diferentes raíces, la matriz de la aplicación se reduce, en un sistema de coordenadas adecuado, a la forma diagonal.*

Por ejemplo, el polinomio característico de la matriz

$$A = \begin{bmatrix} 1 & 3 & 1 & 2 \\ & -1 & 1 & 3 \\ & & 2 & 5 \\ & & & -2 \end{bmatrix}$$

tiene las raíces ± 1, ± 2; las filas de coordenadas de los vectores propios correspondientes son [2, 3, — 5, — 4], [0, 3, — 1, 4], [0, 0, 4, 5] y [0, 0, 0, 1]. Tomándolos como vectores coordenados

reduciremos la matriz A a la forma diagonal con los números 1, -1, 2 y -2 a lo largo de la diagonal principal.

12.2. Células de Jordan. Una matriz de tipo

$$A = \begin{bmatrix} \rho & 1 & 0 & \dots & 0 \\ & \rho & 1 & \dots & 0 \\ & & \rho & \dots & 0 \\ & & & \dots & \dots \\ & & & & \rho \end{bmatrix} \quad (1)$$

se llama *célula de Jordan*. El polinomio característico de una célula de Jordan es igual a $(\lambda-\rho)^n$, donde n es el orden de la matriz. Por consiguiente, ρ es su único valor propio de multiplicidad n.

Sea $\mathfrak{L}$ un espacio lineal de base $e_1, e_2, \dots, e_n$ y sea $\mathcal{A}$ una aplicación lineal que en esta base tiene una matriz A de tipo (1). En este caso tenemos

$$e_1\mathcal{A} = \rho e_1 + e_2, \dots, e_{n-1}\mathcal{A} = \rho e_{n-1} + e_n, e_n\mathcal{A} = \rho e_n \quad (2)$$

y, por consiguiente,

$$e_1(\mathcal{A}-\rho\mathcal{E}) = e_2, e_1(\mathcal{A}-\rho\mathcal{E})^2 = e_3, \dots, e_1(\mathcal{A}-\rho\mathcal{E})^{n-1} = e_n. \quad (3)$$

Puesto que $\mathcal{A}$ debe ser una raíz de su polinomio característico, se tiene $(\mathcal{A}-\rho\mathcal{E})^n = \mathcal{O}$. El polinomio mínimo de la aplicación $\mathcal{A}$ divide su polinomio característico y, por ello, debe ser de la forma $(\lambda-\rho)^s$, $0 < s \leqslant n$. La última de las igualdades (3) muestra que $(\mathcal{A}-\rho\mathcal{E})^{n-1} \neq \mathcal{O}$, de manera que $s = n$, es decir, *el polinomio mínimo de una célula de Jordan coincide con su polinomio característico* $(\lambda-\rho)^n$.

Indiquemos por $\mathfrak{L}_i$ el subespacio tendido sobre los vectores $e_i, e_{i+1}, \dots, e_n$ $(i = 1, 2, \dots, n)$ de la base. De las igualdades (2) y de la forma de la matriz A se deduce que todos estos subespacios son invariantes. Empleando las relaciones (3) es fácil comprobar que $\mathfrak{L}_i$ está compuesto por aquellos vectores, y sólo aquéllos, x para los cuales se tiene

$$x(\mathcal{A}-\rho\mathcal{E})^{n-i+1} = o.$$

Esto demuestra que la cadena de subespacios $\mathfrak{L} = \mathfrak{L}_1 \supset \mathfrak{L}_2 \supset \dots \dots \supset \mathfrak{L}_n \supset o$ no depende de la selección del sistema de coordenadas y está definida por la propia aplicación $\mathcal{A}$.

Demostremos que la aplicación no tiene otros subespacios invariantes. En efecto, sea $\mathfrak{M}$ un subespacio invariante de $\mathcal{A}$ diferente de $\mathfrak{L}_i$. Busquemos un índice i tal que $\mathfrak{M} \supset \mathfrak{L}_i$ y $\mathfrak{M} \not\supset \mathfrak{L}_{i-1}$, aceptando, para generalizar, que $\mathfrak{L}_{n+1} = o$. Mostremos que $\mathfrak{M} = \mathfrak{L}_i$. Consideremos, para ello, un vector arbitrario

$$a = \alpha_i e_i + \alpha_{i+1} e_{i+1} + \dots + \alpha_n e_n \quad (\alpha_i \neq 0) \quad (4)$$

de $\mathfrak{M}$. Para $j \geqslant i$ tenemos $a \in \mathfrak{L}_i$. Supongamos que $j < i$. Aplicando a ambos miembros de (4) la aplicación $(\mathcal{A} - \rho\mathcal{E})^{i-j-1}$, obtenemos

$$a(\mathcal{A} - \rho\mathcal{E})^{i-j-1} = \alpha_j e_{i-1} + \alpha_{j+1} e_i + \ldots + \alpha_{n-i+j+1} e_n \in \mathfrak{M}.$$

Puesto que, por hipótesis, $e_i, \ldots, e_n \in \mathfrak{M}$, resulta que $e_{i-1} \in \mathfrak{M}$ y, por consiguiente, $\mathfrak{L}_{i-1} \subset \mathfrak{M}$ lo que contradice a la elección de i. Por esto tenemos $\mathfrak{M} \subset \mathfrak{L}_i$, de donde $\mathfrak{M} = \mathfrak{L}_i$.

Observemos además que *la matriz de la aplicación $\mathcal{A}$ no se descompone en ningún sistema de coordenadas.*

Efectivamente, la descomposición de la matriz de $\mathcal{A}$ equivale a la descomposición de $\mathfrak{L}$ en una suma directa de subespacios invariantes, lo que es imposible ya que uno de dos subespacios invariantes cualesquiera de la aplicación $\mathcal{A}$ debe estar contenido en el otro, mientras que los sumandos de una suma directa tienen intersección nula.

12.3. Subespacios radicales. Las aplicaciones lineales, cuyas matrices pueden ser reducidas a la forma diagonal o a una célula de Jordan, no abarcan todo el conjunto de matrices. En el caso general, la matriz de cualquier aplicación lineal sobre el cuerpo de los números complejos puede ser reducida a la forma celular diagonal con células de Jordan a lo largo de la diagonal. Se dice que las matrices de este último tipo son *de Jordan* o *que tienen la forma normal de Jordan.*

Supongamos que en una base $e_1, \ldots, e_n$ la matriz de una aplicación lineal $\mathcal{A}$ es de la forma normal de Jordan

$$A = A_1 + A_2 + \ldots + A_s, \tag{5}$$

donde A_i es la célula de Jordan de orden n_i con el valor propio ρ_i $(i = 1, \ldots, s)$ y $n_1 + \ldots + n_s = n$. A la célula A_i le corresponde el subespacio invariante $\mathfrak{L}^{(i)}$ tendido sobre los vectores $e_{p_i+1}, e_{p_i+2}, \ldots, e_{q_i}$ $(p_i = n_1 + \ldots + n_{i-1}, q_i = p_i + n_i)$. La aplicación $\mathcal{A}$ induce en el subespacio $\mathfrak{L}^{(i)}$ una aplicación $\mathcal{A}_i$, cuya matriz es precisamente la célula A_i. Según lo expuesto, todos los vectores x de $\mathfrak{L}^{(i)}$ satisfacen la relación

$$x(\mathcal{A}_i - \rho_i \mathcal{E}_i)^{n_i} = 0$$

y, por consiguiente, también la relación

$$x(\mathcal{A} - \rho_i \mathcal{E})^{n_i} = 0. \tag{6}$$

Sin embargo, ahora ya no se puede afirmar que la relación (6) caracteriza sólo los vectores de $\mathfrak{L}^{(i)}$, ya que entre las células diagonales pueden aparecer otras células con el mismo valor propio. Con el fin de examinar este problema más detalladamente, introduciremos la definición siguiente.

Un vector a se llama *vector radical de altitud h correspondiente a la raíz ρ de una aplicación $\mathcal{A}$*, si

$$a(\rho\mathcal{E}-\mathcal{A})^h=o.$$

El concepto de vector radical es una generalización del concepto de vector propio, ya que los vectores propios son vectores radicales de altitud 1.

El conjunto de todos los vectores radicales correspondientes a una raíz fija ρ de una aplicación $\mathcal{A}$ es un subespacio invariante $\mathfrak{L}_\rho$ llamado subespacio radical de la aplicación $\mathcal{A}$.

Efectivamente, si x e y pertenecen a $\mathfrak{L}_\rho$ y son de altitud h_1 y h_2, tenemos para $h=\max(h_1, h_2)$

$$(\alpha x+\beta y)(\rho\mathcal{E}-\mathcal{A})^h=\alpha x(\rho\mathcal{E}-\mathcal{A})^h+\beta y(\rho\mathcal{E}-\mathcal{A})^h=o,$$
$$x\mathcal{A}(\rho\mathcal{E}-\mathcal{A})^h=x(\rho\mathcal{E}-\mathcal{A})^h\mathcal{A}=o.$$

Los vectores radicales correspondientes a diferentes raíces son necesariamente linealmente independientes. Es más, tiene lugar un teorema más general.

TEOREMA 3. *Si una suma $x_1+x_2+\ldots+x_m=x$ de vectores radicales correspondientes a diferentes raíces $\rho_1, \ldots, \rho_m$ de la aplicación $\mathcal{A}$ está contenida en un subespacio invariante $\mathfrak{M}$, todo sumando por separado está contenido en $\mathfrak{M}$.*

Pongamos

$$\varphi(\lambda)=(\lambda-\rho_1)^{h_1}(\lambda-\rho_2)^{h_2}\ldots(\lambda-\rho_{m-1})^{h_{m-1}}.$$

Por hipótesis, $x\varphi(\mathcal{A})\in\mathfrak{M}$ y, al mismo tiempo,

$$x_1\varphi(\mathcal{A})=x_2\varphi(\mathcal{A})=\ldots=x_{m-1}\varphi(\mathcal{A})=o.$$

Por consiguiente, $x_m\varphi(\mathcal{A})\in\mathfrak{M}$. Los polinomios $\varphi(\lambda)$ y $(\lambda-\rho_m)^{h_m}$ son primos entre sí. Luego, existen unos polinomios $F(\lambda)$ y $G(\lambda)$ tales que

$$1=\varphi(\lambda)F(\lambda)+(\lambda-\rho_m)^{h_m}G(\lambda),$$

de donde

$$\mathcal{E}=\varphi(\mathcal{A})F(\mathcal{A})+(\mathcal{A}-\rho_m\mathcal{E})^{h_m}G(\mathcal{A})$$

y, por consiguiente,

$$x_m=x_m\varphi(\mathcal{A})F(\mathcal{A})+x_m(\mathcal{A}-\rho_m\mathcal{E})^{h_m}G(\mathcal{A})=x_m\varphi(\mathcal{A})F(\mathcal{A})\in\mathfrak{M}$$

que es lo que se quería demostrar.

La afirmación expuesta anteriormente acerca de la independencia lineal de los vectores $x_1, \ldots, x_m$ se obtiene del teorema demostrado tomando $\mathfrak{M}=o$. Como corolario notemos también que *diferentes subespacios radicales tienen intersección nula.*

Volvamos ahora al caso en el que la matriz de una aplicación $\mathcal{A}$ tiene en una base la forma normal de Jordan (5). Hemos definido más arriba dos series de subespacios: los subespacios radicales $\mathfrak{L}_{\rho_1}, \ldots, \mathfrak{L}_{\rho_m}$ y los subespacios $\mathfrak{L}^{(1)}, \ldots, \mathfrak{L}^{(s)}$ correspondientes a las

células diagonales de la matriz A. Para explicar la relación que existe entre ambas series indiquemos por $\mathfrak{M}^{(1)}$ la suma de aquellos subespacios $\mathfrak{L}^{(i)}$ que corresponden a las células con el valor propio ρ_1 y definamos análogamente $\mathfrak{M}^{(2)}, \ldots, \mathfrak{M}^{(m)}$. En virtud de ello, uniendo en la descomposición

$$\mathfrak{L} = \mathfrak{L}^{(1)} + \mathfrak{L}^{(2)} + \ldots + \mathfrak{L}^{(s)}$$

determinados sumandos obtenemos la descomposición

$$\mathfrak{L} = \mathfrak{M}^{(1)} + \mathfrak{M}^{(2)} + \ldots + \mathfrak{M}^{(m)}. \tag{7}$$

Está claro que $\mathfrak{M}^{(i)} \subseteq \mathfrak{L}_{\rho_i}$ $(i = 1, \ldots, m)$. Por esto, además de (7), tiene lugar también la descomposición

$$\mathfrak{L} = \mathfrak{L}_{\rho_1} + \mathfrak{L}_{\rho_2} + \ldots + \mathfrak{L}_{\rho_m}. \tag{8}$$

Puesto que, en virtud del teorema 3, la suma (8) es directa, obtenemos, comparándola con (7), las igualdades requeridas $\mathfrak{M}^{(i)} = \mathfrak{L}_{\rho_i}$.

Es decir, *si la matriz de una aplicación $\mathcal{A}$ puede ser reducida a la forma normal de Jordan, el espacio $\mathfrak{L}$ es la suma directa de los subespacios radicales de $\mathcal{A}$ y, además, cada uno de los subespacios radicales es, a su vez, la suma directa de los subespacios correspondientes a las células de Jordan con el valor propio dado.*

Los razonamientos expuestos permiten ver también que los subespacios radicales se determinan univocamente por la propia aplicación $\mathcal{A}$ y no dependen de la selección de la base de coordenadas $e_1, \ldots, e_n$. En cuanto a los subespacios $\mathfrak{L}^{(i)}$, ellos dependen, en general, tanto de $\mathcal{A}$ como de la forma de reducción de la matriz a la forma diagonal.

Ejemplos y problemas

1. Supongamos que la aplicación $\mathcal{A}$ tiene en la base $e_1, \ldots, e_6$ la matriz $A = A_1 + A_2 + A_3$, donde $A_1 = A_2 = \begin{bmatrix} 6 & 1 \\ 0 & 6 \end{bmatrix}$ y $A_3 = \begin{bmatrix} 7 & 1 \\ 0 & 7 \end{bmatrix}$. Los subespacios radicales de $\mathcal{A}$ son $\mathfrak{L}_6 = Ke_1 + Ke_2 + Ke_3 + Ke_4$ y $\mathfrak{L}_7 = Ke_5 + Ke_6$ (K es el cuerpo principal), mientras que $\mathfrak{L}^{(1)} = Ke_1 + Ke_2$ y $\mathfrak{L}^{(2)} = Ke_3 + Ke_4$. En la base nueva

$$e_1' = e_1 + e_2,\ e_2' = e_2 + e_4,\ e_3' = e_1 - e_3,\ e_4' = e_2 - e_4,\ e_5' = e_5 \text{ y } e_6' = e_6$$

la matriz de $\mathcal{A}$ será la misma; sin embargo, los subespacios $Ke_1' + Ke_2'$ y $Ke_3' + Ke_4'$ correspondientes a la células de Jordan serán distintos.

2. Hállense los polinomios mínimos de las matrices

$$\begin{bmatrix} 2 & 1 & 1 \\ & 2 & 1 \\ & & 2 \end{bmatrix}, \quad \begin{bmatrix} 2 & 1 & 1 \\ & 3 & 1 \\ & & 4 \end{bmatrix}, \quad \begin{bmatrix} 2 & 1 & 1 \\ & 2 & 1 \\ & & 3 \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} 2 & 1 & 1 \\ & 2 & 0 \\ & & 2 \end{bmatrix}.$$

3. Una matriz sobre el cuerpo de los números complejos puede ser reducida a la forma diagonal si, y sólo si, su polinomio mínimo no tiene raíces múltiples.

4. Si la matriz de una aplicación puede ser reducida a la forma normal de Jordan, todo subespacio invariante suyo es suma directa de sus intersecciones con todos los subespacios radicales de la aplicación.

5. Si una matriz de orden n tiene n *diferentes* números característicos, la aplicación lineal correspondiente tiene un total de 2^n subespacios invariantes incluyendo el subespacio nulo y todo el espacio.

Capítulo IV Matrices polinomiales

Los elementos de casi todas las matrices que hemos estudiado hasta el momento eran números de un cuerpo principal K. Sin embargo, al introducir el concepto de polinomio característico nos hemos visto obligados a considerar la matriz característica $\lambda E - A$, cuyos elementos no son números de K, sino *polinomios* en λ con coeficientes de K, suponiendo, además, que K es un *cuerpo conmutativo*. En el capítulo presente nos ocuparemos del estudio sistemático de las propiedades de las matrices polinomiales, es decir, de las matrices, cuyos elementos son polinomios en λ con coeficientes de un cuerpo conmutativo principal. Aplicaremos después estos resultados al problema consistente en hallar la forma de Jordan de la matriz de una aplicación lineal.

§ 13. Factores invariantes

13.1. Equivalencia. Consideremos una matriz cuadrada de orden n

$$\begin{bmatrix} f_{11}(\lambda) & \dots & f_{1n}(\lambda) \\ \vdots & & \vdots \\ f_{n1}(\lambda) & \dots & f_{nn}(\lambda) \end{bmatrix},$$

cuyos elementos son polinomios en la letra λ con coeficientes de un campo principal K. Llamaremos las matrices de este tipo *polinomiales* o *λ-matrices*[1]. Frecuentemente resulta necesario realizar con λ-matrices las transformaciones siguientes:

I. *La multiplicación de una de las filas por un número de K diferente de cero.*

II. *La adición a una de las filas de la matriz de otra fila multiplicada por un polinomio arbitrario $f(\lambda)$.*

[1] Suponemos en todo momento que las matrices consideradas son cuadradas, aunque muchos de los resultados pueden ser extendidos directamente al caso de λ-matrices rectangulares.

III. *La multiplicación de una de las columnas por un número de K diferente de cero.*

IV. *La adición a una de las columnas de los elementos de otra columna multiplicados por un polinomio arbitrario* $f(\lambda)$.

Estas transformaciones se llaman *transformaciones elementales de* λ*-matrices.* Si con una λ-matriz se realiza una transformación elemental se obtiene de nuevo una λ-matriz; con esta matriz se puede realizar otra transformación elemental, etc. Se dice que una λ-matriz F es *equivalente* a una λ-matriz G, si F se puede obtener de G mediante una cadena de transformaciones elementales. En muchas ocasiones resulta útil el lema siguiente:

Mediante las transformaciones elementales I *y* II *se pueden cambiar entre sí dos filas cualesquiera de una* λ*-matriz y mediante las transformaciones elementales* III *y* IV *se pueden cambiar entre sí dos columnas cualesquiera de la misma.*

En efecto, supongamos que es necesario cambiar entre sí la i-ésima y la j-ésima columnas de una matriz. Es fácil ver que esto se consigue realizando las siguientes transformaciones elementales: 1) agregamos a la i-ésima fila la j-ésima; 2) a la j-ésima fila de la matriz nueva agregamos su i-ésima fila multiplicada por -1; 3) multiplicamos la j-ésima fila de la matriz obtenida por -1 y 4) agregamos a la i-ésima fila de la última matriz su j-ésima fila multiplicada por -1. Si realizamos transformaciones análogas con las columnas, lograremos cambiar de posición la i-ésima y la j-ésima columnas. Hemos demostrado el lema.

De este lema se deduce que *si la matriz* F *difiere de* G *en el orden de las columnas o de las filas, la matriz* F *es equivalente a* G.

De la definición de equivalencia de λ-matrices se desprenden directamente las propiedades siguientes:

1. *La relación de equivalencia es transitiva:* si F es equivalente a G y G es equivalente a H, resulta que F es equivalente a H.

En efecto, G se puede obtener, por hipótesis, de H y F de G mediante una cadena de transformaciones elementales; por consiguiente, F se puede obtener mediante una cadena de transformaciones elementales de H.

2. *La relación de equivalencia es simétrica:* si F es equivalente a G, G es equivalente a F. En otras palabras, si F se puede obtener de G mediante una cadena de transformaciones elementales, también G se puede obtener de F mediante una cadena de transformaciones elementales.

Demostremos primero que si F se puede obtener de G mediante una transformación elemental, también G se puede obtener de F mediante una transformación elemental. Consideremos para ello, uno a uno, los cuatro tipos de las transformaciones elementales. Supongamos que F se obtiene de G mediante la transformación de tipo 1, es decir, multiplicando una fila i-ésima de G por un número $\alpha \neq 0$.

Entonces, multiplicando la i-ésima fila de F por α^{-1}, obtendremos, evidentemente, G. Supongamos ahora que F se obtiene de G mediante una transformación de tipo II, por ejemplo, agregando a la i-ésima fila de la matriz G su j-ésima fila multiplicada por $f(\lambda)$. En este caso, agregando a la i-ésima fila de la matriz F su j-ésima fila multiplicada por $-f(\lambda)$, obtendremos de nuevo G. Lo mismo se puede decir acerca de las transformaciones de tipo III y IV. Vemos, por consiguiente, que para toda transformación elemental existe la transformación elemental inversa que anula el resultado de la primera. Por esto, si la matriz F se obtiene de G mediante una cadena de transformaciones elementales, resulta que realizando las transformaciones inversas en el orden contrario podremos obtener de la matriz F la matriz G que es lo que se quería demostrar.

3. *La relación de equivalencia es reflexiva:* toda matriz es equivalente a sí misma.

Por ejemplo, realizando con F dos transformaciones recíprocamente inversas, obtendremos de nuevo F.

13.2. Forma diagonal. Acabamos de demostrar que la relación de equivalencia es transitiva, simétrica y reflexiva. De aquí se deduce que las λ-matrices se descomponen en clases de matrices equivalentes. Surge el problema: ¿puede indicarse una forma para las λ-matrices tal que en cada una de estas clases haya una matriz de la forma dada, y sólo una? Las formas con esta propiedad se denominan *canónicas*. Demostraremos que para las λ-matrices la forma diagonal, con algunas condiciones complementarias de divisibilidad, es canónica en este sentido.

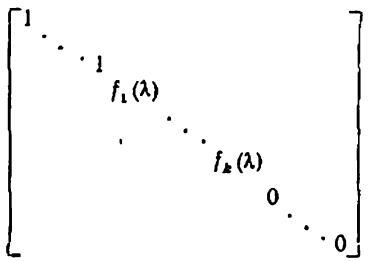
DEFINICIÓN. *Una λ-matriz de tipo*

$$\begin{bmatrix} f_1(\lambda) & & & \\ & f_2(\lambda) & & \\ & & \ddots & \\ & & & f_n(\lambda) \end{bmatrix}$$

se llama canónica diagonal, si todo elemento diagonal $f_i(\lambda)$ divide al elemento siguiente $f_{i+1}(\lambda)$ y si el coeficiente principal de todos los polinomios $f_1(\lambda), \ldots, f_n(\lambda)$ diferentes de cero es 1.

De aquí se deduce que si entre los elementos diagonales de una matriz canónica diagonal aparecen ceros, estos elementos deben ocupar las posiciones últimas, ya que el cero no puede dividir a ningún polinomio no nulo. Por otro lado, si entre los elementos diagonales figuran *números* diferentes de cero, éstos deben ser iguales a 1 y deben ocupar las posiciones primeras, ya que 1 no es divisible por ningún polinomio con coeficiente principal 1, a excepción del polinomio 1. Por consiguiente, en el caso más general la

matriz canónica diagonal es de la forma



donde $f_1(\lambda), \ldots, f_k(\lambda)$ no son constantes, sus coeficientes principales son iguales a 1 y cada uno de ellos es divisor del siguiente.

TEOREMA 1. *Toda λ-matriz puede ser reducida mediante un número finito de tranformaciones elementales a la forma canónica diagonal.*

DEMOSTRACIÓN. Sea G una λ-matriz dada. Si $G = O$, nada hay que demostrar, puesto que O es ya de la forma normal. Supongamos por lo tanto que $G \neq O$. Escojamos entre todas las λ-matrices equivalentes a G aquella cuyo elemento del ángulo superior izquierdo sea diferente de cero y sea del grado mínimo. Sea

$$F = \begin{bmatrix} f_{11}(\lambda) & \ldots & f_{1n}(\lambda) \\ \vdots & & \vdots \\ f_{n1}(\lambda) & \ldots & f_{nn}(\lambda) \end{bmatrix}$$

esta matriz. Demostremos que $f_{11}(\lambda)$ divide a todos los elementos de la primera fila y a todos los elementos de la primera columna de la matriz F. En efecto, sea

$$f_{1i}(\lambda) = f_{11}(\lambda)\, q_i(\lambda) + r_i(\lambda) \quad (i = 1, 2, \ldots, n), \tag{1}$$

donde $q_i(\lambda)$ y $r_i(\lambda)$ son el cociente y el resto de la división de $f_{1i}(\lambda)$ por $f_{11}(\lambda)$. Realicemos con F la siguiente transformación elemental: restemos de los elementos de la i-ésima columna los elementos de la primera multiplicados por $q_i(\lambda)$. La igualdad (1) muestra que el elemento de la primera fila y de la i-ésima columna de la matriz nueva resultará igual a $r_i(\lambda)$. Si $r_i(\lambda) \neq 0$, el grado de $r_i(\lambda)$ es inferior al grado del divisor $f_{11}(\lambda)$. Cambiando entre sí la primera y la i-ésima columnas, obtendremos una matriz equivalente a F en cuyo ángulo superior izquierdo figura un polinomio $r_i(\lambda)$ de grado menor que el de $f_{11}(\lambda)$; pero esto contradice a la definición de la matriz F. Por consiguiente, $r_i(\lambda) = 0$.

Indicando ahora por $q_i(\lambda)$ el cociente de la división de $f_{1i}(\lambda)$ por $f_{11}(\lambda)$, realicemos con la matriz F las siguientes transformaciones elementales: de la segunda columna restemos la primera multiplicada por $q_2(\lambda)$, restemos después de la tercera columna la pri-

mera multiplicada por $q_2(\lambda)$, etc. Realicemos a continuación transformaciones análogas con las filas. La matriz F resultará sustituida por la matriz equivalente

$$H = \begin{bmatrix} f_{11}(\lambda) & 0 & \dots & 0 \\ 0 & h_{22}(\lambda) & \dots & h_{2n}(\lambda) \\ \vdots & \vdots & & \vdots \\ 0 & h_{n2}(\lambda) & \dots & h_{nn}(\lambda) \end{bmatrix}, \qquad (2)$$

donde $h_{ij}(\lambda)$ son unos polinomios determinados.

Todos los polinomios $h_{ij}(\lambda)$ *son divisibles por* $f_{11}(\lambda)$. Efectivamente, si $f_{11}(\lambda)$ no divide a uno de ellos, digamos a $h_{ij}(\lambda)$, entonces sumando a la primera fila de la matriz H su i-ésima fila obtendremos una matriz Q con las siguientes propiedades:

1) Q es equivalente a G,

2) el elemento superior de la izquierda de la matriz Q es diferente de cero y es del menor grado,

3) en la primera fila de la matriz Q figura el elemento $h_{ij}(\lambda)$ que no es divisible por el primer elemento de esta fila.

Sin embargo, hemos visto que de las dos propiedades primeras se desprende que todos los elementos de la primera fila son divisibles por su primer elemento. Por consiguiente, la tercera propiedad contradice a las dos primeras y nuestra proposición queda demostrada. Hemos probado, pues, que para toda λ-matriz G existe una matriz equivalente H de tipo (2), donde todos los $h_{ij}(\lambda)$ son divisibles por $f_{11}(\lambda)$. Realicemos ahora transformaciones elementales con la matriz

$$H_1 = \begin{bmatrix} h_{22}(\lambda) & \dots & h_{2n}(\lambda) \\ \vdots & & \vdots \\ h_{n2}(\lambda) & \dots & h_{nn}(\lambda) \end{bmatrix}.$$

Toda transformación elemental de H_1 puede ser considerada también como una transformación elemental de la matriz H. Es fácil ver que la primera fila y la primera columna de la matriz H no varían en estas transformaciones. Además, puesto que todos los elementos de la matriz H_1 son divisibles por $f_{11}(\lambda)$, todos los elementos de las matrices nuevas, que surgen de H_1 como resultado de transformaciones elementales, también serán divisibles por $f_{11}(\lambda)$.

Aplicando a la matriz H_1 el resultado demostrado anteriormente, veremos que H_1 puede ser reducida mediante transformaciones elementales a la forma

$$\begin{bmatrix} h_{22}(\lambda) & 0 & \dots & 0 \\ 0 & h_{23}(\lambda) & \dots & h_{2n}(\lambda) \\ \vdots & \vdots & & \vdots \\ 0 & h_{n2}(\lambda) & \dots & h_{nn}(\lambda) \end{bmatrix}$$

y, por consiguiente, la matriz H a la forma

$$\begin{bmatrix} f_{11}(\lambda) & & & & \\ & h_{22}(\lambda) & & & \\ & & h_{33}(\lambda) & \dots & h_{3n}(\lambda) \\ & & \vdots & & \vdots \\ & & h_{n3}(\lambda) & \dots & h_{nn}(\lambda) \end{bmatrix},$$

donde todos los $h_{ij}(\lambda)$ $(i, j = 3, \dots, n)$ son divisibles por $h_{22}(\lambda)$ y $h_{22}(\lambda)$ es divisible por $f_{11}(\lambda)$. Continuando este proceso obtendremos al cabo de un número finito de pasos la forma canónica diagonal requerida.

De nuestra demostración se puede extraer fácilmente un método práctico de reducción de λ-matrices a la forma canónica diagonal. Su idea consiste en disminuir primero, empleando las transformaciones elementales, el grado del elemento que figura en la primera fila y en la primera columna y en convertir en cero los demás elementos de las mismas. Después de haber logrado este primer objetivo, aplicamos el mismo método al ángulo H_1 que queda, etc.

El teorema 1 afirma que *toda clase de matrices equivalentes contiene al menos una matriz de forma canónica diagonal.* En el punto siguiente quedará demostrado que esta matriz es única en cada clase.

13.3. Máximos comunes divisores de menores. Sea F una λ-matriz de orden n. Consideremos todos sus menores de orden k $(k = 1, 2, \dots, n)$. Estos menores son unos polinomios en λ. Indiquemos por $D_k(\lambda)$ su máximo común divisor [1]. Si resulta que todos los menores de un orden k son iguales a cero, aceptaremos por definición que $D_k(\lambda) = 0$. En particular, $D_1(\lambda)$ es el máximo común divisor de los elementos de la matriz F, mientras que $D_n(\lambda)$ es igual al determinante de F, dividido por su coeficiente principal.

TEOREMA 2 *Las λ-matrices equivalentes tienen un mismo máximo común divisor de los menores de orden k $(k = 1, 2, \dots, n)$.*

Sean F_1 y F_2 dos λ-matrices equivalentes. Indiquemos los máximos comunes divisores de sus menores de orden k por $D_{k1}(\lambda)$ y $D_{k2}(\lambda)$, respectivamente. Debemos probar que $D_{k1}(\lambda) = D_{k2}(\lambda)$. Sabemos que F_2 puede ser obtenida de F_1 mediante una cadena de transformaciones elementales. Supongamos primero que esta cadena consta sólo de una transformación elemental. Por ejemplo, supongamos que F_2 se obtiene de F_1 multiplicando todos los elementos de la i-ésima fila de la matriz F_1 por un número $\alpha \neq 0$. En este caso los menores respectivos de F_1 y de F_2 o bien son idénticos o bien difieren en el

[1] Convendremos en que el máximo común divisor es siempre el común divisor del grado mayor de *coeficiente principal* igual a 1. Por ello, todos los polinomios $D_k(\lambda)$ diferentes de cero son de coeficiente principal igual a 1.

factor constante α. Pero un factor constante no influye en el cálculo del máximo común divisor, de modo que $D_{k1} = D_{k2}$. Lo mismo sucederá en el caso en que F_2 se obtenga de F_1 multiplicando por α los elementos de una de las columnas de la matriz F_1. Supongamos ahora que F_2 se obtiene de F_1 mediante las transformaciones de tipo II o IV; por ejemplo, aceptemos que F_2 se obtiene al agregar a la i-ésima fila de la matriz F_1 los elementos de la j-ésima fila multiplicados por $f(\lambda)$. Demostremos que D_{k2} es divisible por D_{k1}.

En efecto, existen tres clases de menores de orden k de las matrices F_1 y F_2. A la primera pertenecen aquellos que no contienen elementos de la i-ésima fila. En este caso los menores respectivos de las matrices F_1 y F_2 son, obviamente, iguales. A la segunda clase pertenecen aquellos menores que contienen elementos de la i-ésima y de la j-ésima filas. Estos menores de las matrices F_1 y F_2 también serán iguales, ya que el valor de un determinante no se altera si a los elementos de una de sus filas se suman cantidades proporcionales a los elementos de otra de sus filas. Finalmente, a la tercera clase pertenecen los menores que contienen elementos de la i-ésima fila y no contienen elementos de la j-ésima fila. Los menores respectivos de esta clase son de forma

$$M_1 = \begin{vmatrix} \cdots & \cdots & \cdots \\ f_{iv_1} & \cdots & f_{iv_k} \\ \cdots & \cdots & \cdots \end{vmatrix}, \quad M_2 = \begin{vmatrix} \cdots & \cdots & \cdots \\ f_{iv_1} + f_{jv_1} \cdot f(\lambda) & \cdots & f_{iv_k} + f_{jv_k} \cdot f(\lambda) \\ \cdots & \cdots & \cdots \end{vmatrix},$$

donde las filas de ambos menores que no han sido escritas coinciden. En virtud del teorema de adición de determinantes, tenemos

$$M_2 = \begin{vmatrix} \cdots & \cdots & \cdots \\ f_{iv_1} & \cdots & f_{iv_k} \\ \cdots & \cdots & \cdots \end{vmatrix} + f(\lambda) \begin{vmatrix} \cdots & \cdots & \cdots \\ f_{jv_1} & \cdots & f_{jv_k} \\ \cdots & \cdots & \cdots \end{vmatrix} = M_1 + f(\lambda) N_1,$$

donde N_1 es un menor de la matriz F_1. Todos los menores de orden k de la matriz F_1 son divisibles por D_{k1}. De nuestros razonamientos se ve que D_{k1} divide a todos los menores de orden k de la matriz F_2, ya que éstos o bien coinciden con los menores respectivos de la matriz F_1 o bien se expresan linealmente en términos de los mismos. Pero en tal caso, D_{k1} aparecerá como factor en el máximo común divisor de los menores de orden k de la matriz F_2, es decir, será un factor en D_{k2}. Por lo tanto, si F_2 se obtiene de F_1 mediante una sola transformación elemental, resulta que D_{k2} es divisible por D_{k1}. Realizando con F_2 la transformación elemental inversa, obtendremos F_1. Por ello, D_{k1} debe ser también divisible por D_{k2}. Pero, si los coeficientes principales de dos polinomios coinciden y si estos polinomios son divisibles uno por otro, ellos deben coincidir. Es decir, $D_{k1} = D_{k2}$. Por ahora hemos demostrado la igualdad de los máximos comunes divisores suponiendo que F_2 se obtiene de F_1

mediante una sola transformación elemental. Sin embargo, si $D_k(\lambda)$ no varía en cualquier transformación elemental concreta, es obvio que $D_k(\lambda)$ tampoco varía en el caso de varias transformaciones sucesivas. Por lo tanto, podemos considerar que el teorema 2 ha quedado demostrado en su forma general.

Calculemos los polinomios $D_1(\lambda), \ldots, D_n(\lambda)$ de una matriz de forma canónica diagonal

$$D = \begin{bmatrix} d_1(\lambda) & & & \\ & d_2(\lambda) & & \\ & & \ddots & \\ & & & d_n(\lambda) \end{bmatrix}.$$

Para obtener un menor de orden k debemos suprimir $n-k$ filas y $n-k$ columnas de D. Si suprimimos en D la i-ésima fila, en su i-ésima columna quedarán sólo ceros. Por ello, para obtener un menor diferente de cero, debemos suprimir en D todas las columnas cuyos números coinciden con los de las filas suprimidas. Es decir, los menores de orden k diferentes de cero deben ser de la forma

$$\begin{vmatrix} d_{\nu_1}(\lambda) & & & \\ & d_{\nu_2}(\lambda) & & \\ & & \ddots & \\ & & & d_{\nu_k}(\lambda) \end{vmatrix} = d_{\nu_1}(\lambda)\, d_{\nu_2}(\lambda) \ldots d_{\nu_k}(\lambda). \tag{3}$$

El máximo común divisor de estos menores es $D_k(\lambda)$. De las desigualdades $1 \leqslant \nu_1 \leqslant \nu_2 \leqslant \ldots \leqslant \nu_k \leqslant n$ se deduce que $1 \leqslant \nu_1$, $2 \leqslant \nu_2$, $\ldots$, $k \leqslant \nu_k$. Por lo tanto $d_{\nu_i}(\lambda)$ es divisible por $d_i(\lambda)$, de modo que $d_{\nu_1}(\lambda) \ldots d_{\nu_k}(\lambda)$ es divisible por $d_1(\lambda) \ldots d_k(\lambda)$. Vemos, por consiguiente, que todos los menores de orden k de la matriz D son divisibles por el menor

$$\begin{vmatrix} d_1(\lambda) & & & \\ & \ddots & & \\ & & \ddots & \\ & & & d_k(\lambda) \end{vmatrix} = d_1(\lambda) \ldots d_k(\lambda). \tag{4}$$

Si este menor es igual a cero, todos los menores de orden k de la matriz D también son iguales a cero. Por definición, tenemos en este caso $D_k(\lambda) = 0$. Si el menor (4) es diferente de cero, los polinomios $d_1(\lambda), \ldots, d_k(\lambda)$ son diferentes de cero y sus coeficientes principales son iguales a 1. Pero entonces también el coeficiente principal del menor (4) es igual a 1. Puesto que (4) divide a todos los menores

(3), resulta que $D_k(\lambda)$ coincide con (4). En ambos casos tenemos, por consiguiente,

$$D_k(\lambda) = d_1(\lambda)\, d_2(\lambda) \ldots d_k(\lambda) \quad (k = 1, 2, \ldots, n). \tag{5}$$

Así se calculan los polinomios $D_k(\lambda)$ de una matriz canónica diagonal de elementos diagonales $d_1(\lambda), \ldots, d_n(\lambda)$.

Consideremos ahora una λ-matriz arbitraria F. Indiquemos por $D_k(\lambda)$ el máximo común divisor de los menores de orden k de esta matriz. Según el teorema 1, la matriz F puede ser reducida mediante transformaciones elementales a la forma canónica diagonal

$$D = \begin{bmatrix} d_1(\lambda) & & \\ & \ddots & \\ & & d_n(\lambda) \end{bmatrix}.$$

De acuerdo con el teorema 2 los polinomios $D_k(\lambda)$ calculados para la matriz D coinciden con los respectivos polinomios $D_k(\lambda)$ calculados para F. Por consiguiente, *los polinomios $D_k(\lambda)$ de la matriz F y los elementos diagonales de la matriz canónica diagonal D, a la que se puede reducir F, están ligados por las relaciones* (5).

Supongamos que $D_1(\lambda), \ldots, D_r(\lambda)$ son diferentes de cero y que los demás polinomios $D_{r+1}(\lambda), \ldots, D_n(\lambda)$, si es que existen, son iguales a cero. Entonces de (5) obtenemos

$$\begin{aligned}
&D_1(\lambda) = d_1(\lambda), && d_1(\lambda) = D_1(\lambda),\\
&D_2(\lambda) = d_1(\lambda)\, d_2(\lambda), && d_2(\lambda) = D_2(\lambda) : D_1(\lambda),\\
&\cdots\cdots\cdots && \cdots\cdots\cdots\\
&D_r(\lambda) = d_1(\lambda)\, d_2(\lambda) \ldots d_r(\lambda), && d_r(\lambda) = D_r(\lambda) : D_{r-1}(\lambda),\\
&D_{r+1}(\lambda) = d_1(\lambda)\, d_2(\lambda) \ldots d_r(\lambda)\, d_{r+1}(\lambda), && d_{r+1}(\lambda) = D_{r+1}(\lambda) : D_r(\lambda).
\end{aligned}$$

Puesto que $d_{r+1}(\lambda) = 0$, resulta que $d_{r+2}(\lambda), \ldots, d_n(\lambda)$ también deben ser iguales a cero y obtenemos definitivamente

$$d_1(\lambda) = D_1(\lambda),\ d_2(\lambda) = D_2(\lambda) : D_1(\lambda),\ \ldots,\ d_r(\lambda) = D_r(\lambda) : D_{r-1}(\lambda),$$
$$d_{r+1}(\lambda) = \ldots = d_n(\lambda) = 0. \tag{6}$$

De esta forma hemos obtenido el teorema siguiente.

TEOREMA 3. *Si los máximos comunes divisores $D_k(\lambda)$ de los menores de orden k de una λ-matriz F son diferentes de cero para $k = 1, 2, \ldots, r$ y si $D_{r+1}(\lambda) = 0$, los elementos diagonales $d_k(\lambda)$ de la matriz canónica diagonal, a la que se reduce F mediante transformaciones elementales, se expresan en términos de $D_k(\lambda)$ por las fórmulas* (6) *y, por consiguiente, la matriz F los define unívocamente.*

Los polinomios $d_1(\lambda), \ldots, d_n(\lambda)$ se llaman *factores invariantes de la matriz F.*

El número r que figura en las igualdades (6) tiene un sentido muy simple: es el rango de la matriz F. Efectivamente, el rango de la matriz F es el orden máximo de los menores de F diferentes de cero. Si este orden es igual a r, tenemos en consecuencia que $D_r(\lambda) \not\equiv 0$ y que $D_{r+1}(\lambda) \equiv 0$. Recíprocamente, las condiciones $D_r(\lambda) \not\equiv 0$ y $D_{r+1}(\lambda) \equiv 0$ significan que uno de los menores de orden r es diferente de cero y que todos los menores de orden $r+1$ son iguales a cero. Por consiguiente, el rango de F es igual a r.

13.4. Condiciones de equivalencia. Empleando los resultados del punto anterior es fácil encontrar las condiciones que garanticen que dos λ-matrices dadas sean equivalentes. Representaremos estas condiciones en dos formas.

PRIMERA CONDICIÓN DE EQUIVALENCIA. *Para que unas matrices polinomiales de orden n sean equivalentes es necesario y suficiente que los máximos comunes divisores de sus menores de orden k coincidan para $k = 1, 2, \ldots, n$.*

Puesto que la coincidencia de los máximos comunes divisores de los menores equivale a la coincidencia de los factores invariantes respectivos, la primera condición de equivalencia puede ser enunciada del modo siguiente: *para la equivalencia de λ-matrices es necesario y suficiente que sus factores invariantes correspondientes sean iguales.*

La demostración es evidente. En efecto, si dos λ-matrices F y G son equivalentes, sus máximos comunes divisores $D_k(\lambda)$ son idénticos (teorema 2). Viceversa, si los polinomios $D_k(\lambda)$ de F y de G son iguales, las matrices F y G se reducen mediante transformaciones elementales a una misma matriz canónica diagonal (teorema 3). Pero dos matrices equivalentes a una tercera son equivalentes; por consiguiente F es equivalente a G que es lo que se quería demostrar.

SEGUNDA CONDICIÓN DE EQUIVALENCIA. *Para que unas matrices polinomiales F y G de orden n sean equivalentes es necesario y suficiente que satisfagan la relación*

$$G = PFQ,$$

donde P y Q son unas matrices polinomiales de determinantes constantes diferentes de cero[1].

---

[1] Hemos definido la equivalencia de λ-matrices mediante las transformaciones elementales. Con frecuencia el concepto de equivalencia se define también de otra manera. Se dice que dos λ-matrices G y F son equivalentes, si existen unas matrices cuadradas regulares P y Q de determinantes constantes que satisfacen la relación $G = PFQ$. Entonces, la segunda condición de equivalencia puede ser interpretada como el teorema de equivalencia de estos conceptos.

Antes de pasar a la demostración de esta proposición hagamos algunas observaciones. Sea

$$A = \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & \alpha & \\ & & & & 1 \\ & & & & & \ddots \\ & & & & & & 1 \end{bmatrix} \quad i\text{-ésima fila,}$$

donde α es un número diferente de cero. Multiplicando a la izquierda por A una matriz cualquiera F, veremos que todos los elementos de la matriz F permanecerán intactos, menos los elementos de la i-ésima fila que resultarán multiplicados por α. Por consiguiente, *toda transformación elemental de tipo* I *realizada con la matriz F equivale a la multiplicación de F a la izquierda por una matriz A convenientemente escogida.* Análogamente, si multiplicamos la matriz F a la izquierda por la matriz

$$B = \begin{bmatrix} 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \dots & 1 & \dots & f(\lambda) & \dots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & \dots & 1 & \dots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & \dots & 0 & \dots & 1 \end{bmatrix} \begin{matrix} \\ \\ i\text{-ésima fila} \\ \\ j\text{-ésima fila,} \\ \\ \end{matrix}$$

donde todos los elementos diagonales son iguales a la unidad, el elemento de la i-ésima fila y de la j-ésima columna es igual a $f(\lambda)$ y los restantes elementos son iguales a cero, resultará que a los elementos de la i-ésima fila de F se sumarán los elementos de su j-ésima fila multiplicados por $f(\lambda)$. Por consiguiente, *toda transformación elemental de tipo* II *equivale a la multiplicación a la izquierda de la matriz F por una matriz B convenientemente escogida.*

Finalmente, es fácil comprobar de esta misma forma que *las transformaciones elementales de la matriz F de tipo* III *y* IV *equivalen a la multiplicación de F a la derecha por las matrices respectivas A y B.*

Pasemos ahora a demostrar la segunda condición de equivalencia.

NECESIDAD. Sea G una matriz equivalente a la matriz F. Esto significa que G se puede obtener de F mediante una cadena de transformaciones elementales sucesivas. Podemos sustituir cada una de las transformaciones elementales por la multiplicación por una matriz de tipo A o B a la izquierda o a la derecha, respectivamente. Como resultado obtendremos la igualdad

$$G = P_1 P_2 \dots P_p F Q_1 Q_2 \dots Q_q, \tag{7}$$

donde cada una de las matrices P_i y Q_j es de tipo A o B ($i, j = 1, 2, \dots$). Pongamos

$$P = P_1 P_2 \dots P_p \text{ y } Q = Q_1 Q_2 \dots Q_q.$$

Puesto que los determinantes de las matrices B son iguales a la unidad y los determinantes de las matrices A son números constantes diferentes de cero, resulta que los determinantes de las matrices P y Q también serán unos números constantes diferentes de cero. La relación (7) muestra que

$$G = PFQ$$

y la necesidad queda demostrada.

SUFICIENCIA. Supongamos al contrario que

$$G = PFQ, \tag{8}$$

donde P y Q son matrices polinomiales de determinantes constantes diferentes de cero. El máximo común divisor $D_n(\lambda)$ de todos los menores de orden n de la matriz P es igual al determinante de P, dividido por su coeficiente principal. Puesto que este determinante es un número constante, resulta que $D_n(\lambda) = 1$. Para $k = n$ tenemos de las relaciones (5)

$$D_n(\lambda) = d_1(\lambda) d_2(\lambda) \dots d_n(\lambda) = 1,$$

y de aquí resulta

$$d_1(\lambda) = d_2(\lambda) = \dots = d_n(\lambda) = 1,$$

donde $d_1(\lambda), \dots, d_n(\lambda)$ son los factores invariantes de la matriz P. Pero los factores invariantes de la matriz unidad E también son todos iguales a la unidad, ya que E es de forma canónica diagonal. Según el primer criterio de equivalencia, de aquí se deduce que la matriz P es equivalente a E y, por consiguiente, puede ser obtenida de E mediante una cadena de transformaciones elementales. Toda transformación elemental puede ser sustituida por la multiplicación por una matriz de tipo A o B. Así P quedará representada en la forma siguiente

$$P = P_1 \dots P_k E Q_1 \dots Q_l = P_1 \dots P_k Q_1 \dots Q_l,$$

donde P_i y Q_j son matrices de tipo A y B.

Aplicando estos mismos razonamientos a la matriz Q, obtenemos un desarrollo análogo

$$Q = M_1 \ \dots \ M_s N_1 \ \dots \ N_t.$$

Introduciendo en (8) estas descomposiciones, llegamos a la igualdad

$$G = P_1 \ \dots \ P_k Q_1 \ \dots \ Q_t F M_1 \ \dots \ M_s N_1 \ \dots \ N_t \qquad (9)$$

de la cual se ve que G se obtiene multiplicando sucesivamente la matriz F por las matrices P_i, Q_i, M_i y N_i de tipo A o B. Pero cada multiplicación de esta índole equivale a una transformación elemental. Por consiguiente, G es equivalente a F y hemos concluido la demostración.

Ejemplos y problemas

1. Redúzcanse a la forma canónica diagonal mediante transformaciones elementales las matrices

$$\begin{bmatrix} \lambda-2 & -1 & 0 \\ 0 & \lambda-2 & -1 \\ 0 & 0 & \lambda-2 \end{bmatrix}, \quad \begin{bmatrix} \lambda(\lambda+1) & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & (\lambda+1)^2 \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} 1-\lambda & \lambda^2 & \lambda \\ \lambda & \lambda & -\lambda \\ 1+\lambda^2 & \lambda^2 & -\lambda^2 \end{bmatrix}.$$

2. Empleando los máximos comunes divisores de los menores, hállese la forma canónica diagonal de las λ-matrices

$$\begin{bmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \lambda & 1 \\ 5 & 4 & 3 & \lambda+2 \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} \alpha+\lambda & \beta & 1 & 0 \\ -\beta & \alpha+\lambda & 0 & 1 \\ 0 & 0 & \alpha+\lambda & \beta \\ 0 & 0 & -\beta & \alpha+\lambda \end{bmatrix}.$$

3. Demuéstrese que toda λ-matriz rectangular de m filas y de n columnas puede ser reducida mediante transformaciones elementales a la forma

$$\begin{bmatrix} d_1\lambda & 0 & \dots & 0 & \dots & 0 \\ 0 & d_2(\lambda) & \dots & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & d_m(\lambda) & \dots & 0 \end{bmatrix} \quad \text{o a la forma} \quad \begin{bmatrix} d_1(\lambda) & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & d_n(\lambda) \\ 0 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & 0 \end{bmatrix}.$$

¿Cómo debe enunciarse la segunda condición de equivalencia para estas matrices?

4. Demuéstrese que mediante las transformaciones elementales de tipo I y II toda λ-matriz cuadrada puede ser reducida a la forma

$$\begin{bmatrix} f_{11}(\lambda) & f_{12}(\lambda) & \dots & f_{1n}(\lambda) \\ & f_{22}(\lambda) & \dots & f_{2n}(\lambda) \\ & & \dots & \\ & & & f_{nn}(\lambda) \end{bmatrix},$$

donde los elementos diagonales o bien son iguales a cero o bien son de coeficiente principal igual a 1.

5. Consideremos en lugar de λ-matrices las matrices formadas por números enteros. Las transformaciones elementales de estas matrices se definen del modo siguiente: I, multiplicación de una fila por ± 1; II, adición a los elementos de una fila de los elementos de otra cualquiera multiplicados por un número entero; III y IV, transformaciones similares de las columnas. Una matriz diagonal

formada por números enteros se llama canónica, si sus elementos diagonales son no negativos y si todo elemento diagonal anterior divide al que le sigue. Demuéstrense los teoremas: a) toda matriz formada por números enteros puede ser reducida, mediante un número finito de transformaciones elementales, a la forma canónica diagonal; b) esta forma canónica diagonal es única y c) para la equivalencia de dos matrices F y G formadas por números enteros es necesario y suficiente que para ellas sea válida la relación $G = PFQ$, donde P y Q son unas matrices formadas por números enteros, cuyos determinantes son iguales a ± 1.

§ 14. Divisores elementales

Los factores invariantes determinan, salvo una equivalencia, la matriz polinomial F. Sin embargo, si F se descompone en células diagonales, la relación entre los factores invariantes de la matriz F y los factores invariantes de las células resulta compleja. Por ello, conviene considerar en algunas cuestiones, en lugar de los factores invariantes, los así llamados divisores elementales de la matriz F, ya que el comportamiento de estos últimos en el caso de descomposición de F es más sencillo.

14.1. Relación con los factores invariantes. Consideremos una λ-matriz arbitraria F, cuyos elementos son polinomios en λ con coeficientes del campo principal K. No someteremos el cuerpo conmutativo K a ninguna restricción. Sean $d_1(\lambda)$, $d_2(\lambda)$, ..., $d_n(\lambda)$ los factores invariantes de la matriz F. Parte de estos factores puede ser igual a cero; por ello, supondremos, para concretar, que $d_1(\lambda)$, ..., $d_r(\lambda)$ son diferentes de cero y que $d_{r+1}(\lambda) = \ldots = d_n(\lambda) = 0$. El número r, como hemos visto anteriormente, es el rango de la matriz F. Descompongamos cada uno de los polinomios $d_1(\lambda)$, ..., $d_r(\lambda)$ en factores irreducibles sobre el cuerpo conmutativo K. Sea, por ejemplo,

$$d_l(\lambda) = [\varepsilon_1(\lambda)]^{\alpha_1} [\varepsilon_2(\lambda)]^{\alpha_2} \ldots [\varepsilon_k(\lambda)]^{\alpha_k},$$

donde $\varepsilon_1(\lambda)$, ..., $\varepsilon_k(\lambda)$ son distintos polinomios irreducibles de coeficiente principal igual a 1. Las expresiones $[\varepsilon_1(\lambda)]^{\alpha_1}$, ..., $[\varepsilon_k(\lambda)]^{\alpha_k}$ se llaman *divisores elementales* del factor invariante $d_l(\lambda)$. Los divisores elementales de todos los factores invariantes de la matriz F que no sean constantes se llaman *divisores elementales* de esta matriz. Por ejemplo, si los factores invariantes de la matriz F son iguales, respectivamente a 1, λ, $\lambda^2(\lambda+1)$ y $\lambda^2(\lambda+1)^2$, sus divisores elementales serán λ, λ^2, λ^2, $\lambda+1$, $(\lambda+1)^2$. Un divisor elemental de tipo $[\varepsilon(\lambda)]^k$, donde $\varepsilon(\lambda)$ es un polinomio irreducible, se llama *perteneciente* al polinomio $\varepsilon(\lambda)$. En el ejemplo considerado los divisores elementales λ, λ^2 y λ^2 pertenecen a λ, mientras que $\lambda+1$ y $(\lambda+1)^2$ pertenecen a $\lambda+1$.

Tomemos ahora una λ-matriz F y escribamos todos sus divisores elementales. Si uno de los divisores elementales de F figura en varios factores invariantes de F, lo escribiremos tantas veces cuantas

aparece en los factores invariantes. Demostremos que el sistema de divisores elementales obtenido de esta forma determina plenamente todos los factores invariantes de la matriz F diferentes de una constante; si tomamos en consideración, además, el orden y el rango de F, quedarán determinados todos los factores invariantes de la matriz F.

TEOREMA 1. *El orden, el rango y el sistema de divisores elementales de una λ-matriz F determinan plenamente sus factores invariantes y, por consiguiente, determinan F salvo una equivalencia.*

La demostración se aclara fácilmente con el ejemplo siguiente. Supongamos que F es de orden 6 y de rango 4 y que sus divisores elementales son λ, λ^2, λ^2, $\lambda+1$, $(\lambda+1)^2$, $\lambda-1$ y $\lambda-1$. Puesto que el orden de F es 6, la matriz F tiene seis factores invariantes $d_1(\lambda)$, ..., $d_6(\lambda)$. Además, $d_5(\lambda)=d_6(\lambda)=0$, ya que el rango de F debe ser 4. Si descomponemos $d_1(\lambda)$, ..., $d_4(\lambda)$ en factores debemos obtener los siete divisores elementales indicados. Puesto que, sin embargo, $d_4(\lambda)$ es divisible por $d_3(\lambda)$, $d_2(\lambda)$ y $d_1(\lambda)$, resulta que en $d_4(\lambda)$ figuran los divisores elementales de potencia superior pertenecientes a todos los polinomios irreducibles. Luego, $d_4(\lambda)=\lambda^2(\lambda+1)^2(\lambda-1)$. Entre los divisores elementales restantes λ, λ^2, $\lambda+1$ y $\lambda-1$ los de potencia superior deben componer $d_3(\lambda)$; por consiguiente, $d_3(\lambda)=\lambda^2(\lambda+1)(\lambda-1)$. A su vez, los de potencia superior de éntre los que quedan deben formar $d_2(\lambda)$, es decir, $d_2(\lambda)=\lambda$. Puesto que todos los divisores elementales han sido ya distribuidos, resulta que $d_1(\lambda)=1$. Es obvio que este método se puede aplicar en cualquier caso, lo que demuestra el teorema.

Los divisores elementales dependen del campo principal K. Por ejemplo, supongamos que los factores invariantes de una λ-matriz F son iguales a λ^2+1 y $(\lambda^2+1)^2$. Si el campo principal es cuerpo de los números reales, el polinomio λ^2+1 es irreducible y los divisores elementales de la matriz F son λ^2+1 y $(\lambda^2+1)^2$. Sin embargo, si el campo principal es cuerpo de los números complejos, se tiene $\lambda^2+1=(\lambda-i)(\lambda+i)$ y los divisores elementales de F serán $\lambda+i$, $(\lambda+i)^2$, $\lambda-i$ y $(\lambda-i)^2$.

14.2. Divisores elementales de una matriz descompuesta. Para obtener los divisores elementales de una λ-matriz que tiene la forma canónica diagonal es suficiente tomar, de acuerdo con la definición, todos los divisores elementales de sus elementos diagonales. Mostremos que esta misma regla tiene lugar también para una λ-matriz diagonal cualquiera.

LEMA. *El sistema de divisores elementales de una λ-matriz diagonal cualquiera F es la unión de los divisores elementales de sus elementos diagonales.*

Supongamos que los elementos diagonales de F son los polinomios $f_1(\lambda)$, $f_2(\lambda)$, ..., $f_n(\lambda)$. Podemos aceptar, sin perder genera-

lidad, que todos estos polinomios son diferentes de cero y que sus coeficientes principales son iguales a 1. Indiquemos por $D_k(\lambda)$ el máximo común divisor de los menores de orden k de la matriz F. Puesto que los coeficientes principales de los polinomios $f_1(\lambda), \ldots, f_n(\lambda)$ son iguales a la unidad, resulta que $D_n(\lambda)$ es el determinante de la matriz F, es decir,

$$D_n(\lambda) = f_1(\lambda) f_2(\lambda) \ldots f_n(\lambda).$$

Pero

$$D_n(\lambda) = d_1(\lambda) d_2(\lambda) \ldots d_n(\lambda),$$

donde $d_1(\lambda), \ldots, d_n(\lambda)$ son los factores invariantes de la matriz F. Por esto, indicando por $\varepsilon_1(\lambda), \varepsilon_2(\lambda), \ldots, \varepsilon_s(\lambda)$ los distintos factores irreducibles de los polinomios $f_1(\lambda), \ldots, f_n(\lambda)$, podemos ver que todo divisor elemental de la matriz F es una potencia de uno de los polinomios $\varepsilon_1(\lambda), \ldots, \varepsilon_s(\lambda)$. Despejemos ahora en $f_1(\lambda), \ldots, f_n(\lambda)$ las potencias máximas de $\varepsilon_1(\lambda)$ por las que son divisibles estos polinomios. Sea

$$f_i(\lambda) = [\varepsilon_1(\lambda)]^{s_i} g_i(\lambda) \quad (i = 1, 2, \ldots, n),$$

donde $g_i(\lambda)$ no es divisible por $\varepsilon_1(\lambda)$. Queremos probar que $[\varepsilon_1(\lambda)]^{s_1}, \ldots, [\varepsilon_1(\lambda)]^{s_n}$ es precisamente el sistema de divisores elementales de la matriz F pertenecientes al polinomio irreducible $\varepsilon_1(\lambda)$. Puesto que los divisores elementales de la matriz F no dependen del orden de sus filas y de sus columnas, podemos disponer estas filas y columnas de modo que

$$s_1 \leqslant s_2 \leqslant \ldots \leqslant s_n. \tag{1}$$

Hallemos la potencia superior con la que $\varepsilon_1(\lambda)$ figura en $D_k(\lambda)$. Por definición, $D_k(\lambda)$ es el máximo común divisor de los menores de orden k de la matriz F, entre los cuales, como hemos visto anteriormente (p. 13.3), serán diferentes de cero sólo los menores iguales a

$$f_{v_1}(\lambda) \ldots f_{v_k}(\lambda) = [\varepsilon_1(\lambda)]^{s_{v_1} + \ldots + s_{v_k}} g_{v_1}(\lambda) \ldots g_{v_k}(\lambda).$$

En vista de las desigualdades (1), la menor potencia de $\varepsilon_1(\lambda)$ la tiene el menor

$$f_1(\lambda) \ldots f_k(\lambda) = [\varepsilon_1(\lambda)]^{s_1 + \ldots + s_k} g_1(\lambda) \ldots g_k(\lambda).$$

Por consiguiente, $D_k(\lambda)$ contiene $\varepsilon_1(\lambda)$ en la potencia $s_1 + \ldots + s_k$. Sustituyendo aquí k por $k-1$, obtenemos que $D_{k-1}(\lambda)$ contiene $\varepsilon_1(\lambda)$ en la potencia $s_1 + \ldots + s_{k-1}$. Pero el factor invariante $d_k(\lambda)$ es igual al cociente de $D_k(\lambda)$ por $D_{k-1}(\lambda)$. Por esto $d_k(\lambda)$ contiene $\varepsilon_1(\lambda)$ exactamente en la potencia s_k. Luego, los divisores elementales de la matriz F pertenecientes al polinomio irreducible $\varepsilon_1(\lambda)$ coinciden con los divisores elementales pertenecientes a $\varepsilon_1(\lambda)$ de los

elementos diagonales de la matriz F. Puesto que nuestro razonamiento es válido también para los demás polinomios $e_2(\lambda), \ldots, e_r(\lambda)$, el lema queda demostrado en el caso más general.

TEOREMA 2. *El sistema de divisores elementales de una matriz descompuesta es igual a la unión de los sistemas de divisores elementales de sus células.*

Sea F una λ-matriz de forma celular diagonal

$$F = \begin{bmatrix} F_1 & & \\ & \ddots & \\ & & F_m \end{bmatrix}.$$

Las transformaciones elementales de cada una de las células $F_1, \ldots, F_m$ pueden ser consideradas como transformaciones de toda la matriz F. Estas transformaciones no alteran la estructura celular diagonal de la matriz F y las transformaciones, realizadas con una de las células, no alteran la forma de las restantes. Por esto, mediante transformaciones elementales de la matriz F se pueden reducir todas las células a la forma diagonal. Aplicando el lema vemos que los divisores elementales de las matrices $F, F_1, \ldots, F_m$ serán las uniones de los divisores elementales de aquellos elementos diagonales que figuran en estas matrices. En particular, los divisores elementales de la matriz F se obtienen uniendo los divisores elementales de todos los elementos diagonales, es decir, uniendo los divisores elementales de todas las células $F_1, \ldots, F_m$, que es lo que se quería demostrar.

Ejemplos y problemas

1. Hállense los divisores elementales de las matrices siguientes:

$$\begin{bmatrix} \lambda & 1 & & \\ & \lambda & 1 & \\ & & \lambda & 1 \\ & & & \lambda \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 1 & \lambda+2 \\ 0 & 1 & \lambda+2 & \\ 1 & \lambda+2 & & \\ \lambda+2 & & & \end{bmatrix} \text{ y } \begin{bmatrix} 0 & 0 & 0 & \lambda^2 \\ 0 & 0 & \lambda(\lambda-1) & \\ 0 & (\lambda-1)^2 & & \\ \lambda(\lambda-1) & & & \end{bmatrix}.$$

2. Hállense los factores invariantes de las matrices:

$$\begin{bmatrix} \lambda(\lambda+1) & & & \\ & \lambda^2 & & \\ & & (\lambda+1)^2 & \\ & & & \lambda(\lambda-1) \end{bmatrix}, \quad \begin{bmatrix} \lambda & 1 & 2 & 3 \\ & \lambda & 1 & 2 \\ & & \lambda & 1 \\ & & & \lambda \end{bmatrix} \text{ y } \begin{bmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 1 & \lambda & 0 \\ 0 & 0 & 1 & \lambda \end{bmatrix}.$$

3. Hállense los divisores elementales de la matriz:

$$\begin{bmatrix} \lambda^2+2 & \lambda^2+1 & \lambda^2+1 \\ 3 & \lambda^2+1 & 3 \\ \lambda^2+1 & \lambda^2+1 & \lambda^2+1 \end{bmatrix}$$

en el cuerpo de los números racionales, en el cuerpo de los números reales y en el cuerpo de los números complejos.

§ 15. Formas normales de la matriz de una aplicación lineal

15.1. División de λ-matrices. Se llama *polinomio matricial* en la variable λ la expresión de tipo

$$F(\lambda) = A_0\lambda^m + A_1\lambda^{m-1} + A_2\lambda^{m-2} + \ldots + A_m, \qquad (1)$$

donde $A_0, \ldots, A_m$ son matrices cuadradas de un mismo orden y con elementos de un campo principal K. Dos polinomios se llaman iguales, si son iguales las matrices que figuran en estos polinomios en los términos con la misma potencia de la variable λ. Los λ-polinomios matriciales se suman y se multiplican siguiendo las reglas habituales. Está claro que todo λ-polinomio puede ser representado mediante una matriz, cuyos términos son polinomios corrientes en λ, y viceversa. Por ejemplo

$$\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} + \begin{bmatrix} 5 & 6 \\ 7 & -2 \end{bmatrix}\lambda + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\lambda^2 = \begin{bmatrix} \lambda^2 + 5\lambda + 1 & 6\lambda + 2 \\ 7\lambda & \lambda^2 - 2\lambda + 3 \end{bmatrix}.$$

Luego, los λ-polinomios matriciales son simplemente una forma especial de representación de λ-matrices.

Si la matriz A_0 de un λ-polinomio (1) es diferente de la nula, se dice que m es el grado del polinomio matricial. El polinomio $F(\lambda)$ se llama *regular*, si la matriz A_0 es invertible.

Es evidente que *el grado de una suma de λ-polinomios matriciales no sobrepasa el máximo de los grados de los sumandos.* Mediante ejemplos es fácil comprobar que el grado del producto de dos λ-polinomios matriciales puede resultar menor que la suma de los grados de los factores. Sin embargo, *si al menos uno de los dos polinomios que se multiplican es regular, el grado del producto es igual exactamente a la suma de los grados de los factores.*

En efecto, si

$$\begin{aligned} A(\lambda) &= A_0\lambda^m + A_1\lambda^{m-1} + \ldots + A_m, \\ B(\lambda) &= B_0\lambda^p + B_1\lambda^{p-1} + \ldots + B_p \end{aligned} \qquad (2)$$

y si B_0 es invertible, el término principal del producto $A(\lambda) \cdot B(\lambda)$ es igual a $A_0B_0\lambda^{m+p}$ y de $A_0B_0 = O$ resulta que $O = A_0B_0B_0^{-1} = A_0$ lo que contradice a lo supuesto.

TEOREMA 1. *Cualesquiera que sean un λ-polinomio matricial $A(\lambda)$ y un λ-polinomio regular $B(\lambda)$ existen unos λ-polinomios $P(\lambda)$, $S(\lambda)$, $Q(\lambda)$ y $R(\lambda)$ que satisfacen las condiciones*

a) $A(\lambda) = B(\lambda)P(\lambda) + S(\lambda)$ y $A(\lambda) = Q(\lambda)B(\lambda) + R(\lambda)$;

b) *o bien $S(\lambda) = O$ o bien el grado de $S(\lambda)$ es menor que el grado de $B(\lambda)$; o bien $R(\lambda) = O$ o bien el grado de $R(\lambda)$ es menor que el grado de $B(\lambda)$.*

Los polinomios indicados se determinan univocamente por las condiciones a) *y* b) *y se dice que* $P(\lambda)$ *y* $S(\lambda)$ *son el cociente y el resto a la izquierda y que* $Q(\lambda)$ *y* $R(\lambda)$ *son el cociente y el resto a la derecha de la división de* $A(\lambda)$ *por* $B(\lambda)$.

La demostración consistirá en que indicaremos un método que permite encontrar los cocientes y los restos y que, en el caso considerado, coincide con el algoritmo de división de los polinomios corrientes. Supongamos, por ejemplo, que $A(\lambda)$ y $B(\lambda)$ son de forma (2). Si es $m < p$, podemos tomar $P(\lambda) = 0$ y $S(\lambda) = A(\lambda)$. Sea, pues, $m \geqslant p$. Formamos las siguientes diferencias sucesivas

$$S_1(\lambda) = A(\lambda) - A_0 B_0^{-1} \lambda^{m-p} B(\lambda) = C_0^{(1)} \lambda^{m_1} + \ldots,$$
$$S_2(\lambda) = S_1(\lambda) - C_0^{(1)} B_0^{-1} \lambda^{m_1 - p} B(\lambda) = C_0^{(2)} \lambda^{m_2} + \ldots,$$
$$\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots$$
$$S_{t+1}(\lambda) = S_t(\lambda) - C_0^{(t)} B_0^{-1} \lambda^{m_t - p} B(\lambda) = C_0^{(t+1)} \lambda^{m_{t+1}} + \ldots,$$

donde los puntos suspensivos representan términos de grados inferiores y donde $m_{t+1} < p$. Sumando estas igualdades y tomando

$$S(\lambda) = S_{t+1}(\lambda),$$
$$P(\lambda) = A_0 B_0^{-1} \lambda^{m-p} + C_0^{(1)} B_0^{-1} \lambda^{m_1 - p} + \ldots + C_0^{(t)} B_0^{-1} \lambda^{m_t - p},$$

obtenemos la primera de las relaciones a). Análogamente se determinan $Q(\lambda)$ y $R(\lambda)$.

Resta demostrar la unicidad. Pero, si

$$A(\lambda) = B(\lambda) P(\lambda) + S(\lambda) = B(\lambda) P_*(\lambda) + S_*(\lambda),$$

se tiene

$$B(\lambda)[P(\lambda) - P_*(\lambda)] = S_*(\lambda) - S(\lambda).$$

Puesto que el polinomio $B(\lambda)$ es regular, resulta que si la diferencia $P(\lambda) - P_*(\lambda)$ es no nula, el grado del primer miembro es no menor que el grado de $B(\lambda)$, mientras que el grado del segundo miembro es, indudablemente, menor que el grado de $B(\lambda)$. Por esto debe ser $P(\lambda) = P_*(\lambda)$ y $S(\lambda) = S_*(\lambda)$.

15.2. Equivalencia escalar. Según el p. 13.4, dos λ-matrices $A(\lambda)$ y $B(\lambda)$ son equivalentes si, y sólo si, existen unas λ-matrices $U(\lambda)$ y $V(\lambda)$, de determinantes no nulos que no dependen de λ, que satisfacen la relación

$$A(\lambda) = U(\lambda) B(\lambda) V(\lambda). \tag{3}$$

Diremos que la matriz $A(\lambda)$ es *escalarmente equivalente* a la matriz $B(\lambda)$, si existen unas matrices regulares U y V, formadas *por elementos que no dependen de* λ, que satisfacen la relación (3). Las matrices formadas por elementos que no dependen de λ se llamarán *escalares*.

TEOREMA 2. *Si dos λ-polinomios de primer grado $A\lambda + B$ y $C\lambda + D$ son regulares y equivalentes, también son escalarmente equivalentes.*

Por hipótesis tenemos

$$A\lambda + B = U(\lambda)(C\lambda + D)V(\lambda), \tag{4}$$

donde $U(\lambda)$ y $V(\lambda)$ son matrices de determinantes constantes diferentes de cero. Indiquemos por P y S el cociente y el resto a la izquierda de la división de $U(\lambda)$ por $A\lambda + B$ e indiquemos por Q y R el cociente y el resto a la derecha de la división de $V(\lambda)$ por $A\lambda + B$. Es decir,

$$U = (A\lambda + B)P + S \quad \text{y} \quad V = Q(A\lambda + B) + R. \tag{5}$$

Las matrices S y R son escalares, ya que son de grado menor que la unidad. Probemos que

$$A\lambda + B = S(C\lambda + D)R. \tag{6}$$

Efectivamente, multiplicando por U^{-1} ambos miembros de la igualdad (4), sustituyendo V por su expresión (5) y agrupando los términos, obtenemos

$$[U^{-1} - (C\lambda + D)Q](A\lambda + B) = (C\lambda + D)R.$$

Comparando los grados del primero y del segundo miembros veremos que la expresión que figura entre los corchetes debe ser igual a una matriz escalar T; tenemos, pues, que

$$T = U^{-1} - (C\lambda + D)Q \quad \text{y} \quad T(A\lambda + B) = (C\lambda + D)R. \tag{7}$$

De aquí resulta que

$$E = U(C\lambda + D)Q + UT = (A\lambda + B)V^{-1}Q + UT = \\ = (A\lambda + B)V^{-1}Q + [(A\lambda + B)P + S]T,$$

es decir,

$$E = (A\lambda + B)[V^{-1}Q + PT] + ST.$$

Pero el segundo miembro puede ser de grado nulo sólo en el caso en que sea cero la expresión que figura entre los corchetes, de donde

$$E = ST \quad \text{y} \quad T = S^{-1}.$$

Comparando con (7), obtenemos (6), donde S y, por consiguiente, R son matrices escalares invertibles.

15.3. Criterio de semejanza de matrices. Según el p. 3.1, dos matrices A y B sobre un cuerpo conmutativo K se llaman *semejantes* si existe una matriz invertible T, formada por elementos de K, tal que

$$A = T^{-1}BT. \tag{8}$$

En el capítulo III se ha explicado la importancia que tiene el encontrar las condiciones de semejanza de unas matrices dadas. Em-

pleando los resultados obtenidos es fácil indicar las condiciones necesarias y suficientes de semejanza de matrices y resolver, con ello, uno de los problemas principales de la teoría de matrices.

TEOREMA 3. *Para que las matrices A y B, definidas sobre un cuerpo conmutativo arbitrario K, sean semejantes es necesario y suficiente que sus matrices características $\lambda E - A$ y $\lambda E - B$ sean equivalentes.*

La necesidad es evidente, ya que de (8) se deduce que

$$\lambda E - A = T^{-1}(\lambda E - B)\,T,$$

es decir, que $\lambda E - B$ y $\lambda E - A$ son equivalentes y, es más, son escalarmente equivalentes. Viceversa, supongamos que las matrices $\lambda E - B$ y $\lambda E - A$ son equivalentes. Puesto que representan unos λ-polinomios matriciales regulares y de primer grado, resulta que estas matrices son, en virtud del teorema 2, escalarmente equivalentes, es decir, existen unas matrices S y R escalares y regulares tales que

$$\lambda E - A = S(\lambda E - B)\,R.$$

Comparando en esta igualdad los coeficientes de λ y los términos independientes, obtenemos

$$E = SR \quad \text{y} \quad A = SBR,$$

de donde resulta

$$A = R^{-1}BR \tag{9}$$

que es lo que se quería demostrar.

De los teoremas demostrados se puede extraer el siguiente algoritmo para determinar la semejanza de las matrices A y B: formamos las matrices características $\lambda E - A$ y $\lambda E - B$ y, empleando el proceso descrito en el p. 13.2, las reducimos mediante transformaciones elementales a la forma canónica diagonal. Si estas formas coinciden, las matrices A y B son semejantes; si las formas son distintas, las matrices A y B no son semejantes.

A veces, además de establecer el hecho mismo de semejanza, es necesario hallar también la matriz transformadora T tal que $B = T^{-1}AT$. Con este fin, para pequeños valores del orden n de las matrices consideradas, se toma $T = \|t_{ij}\|$, se escribe la igualdad matricial $TB = AT$ en forma de n^2 igualdades entre los elementos de TB y AT y se consideran estas igualdades como ecuaciones lineales homogéneas respecto a las n^2 incógnitas t_{ij} $(i, j = 1, 2, \ldots, n)$. Resolviendo este sistema, se determina T.

El método expuesto resulta muy voluminoso para grandes valores de n y en este caso es preferible seguir el camino indicado en la propia demostración del teorema 3. Ante todo, conociendo las transformaciones elementales que reducen las matrices $\lambda E - A$ y $\lambda E - B$ a la forma canónica diagonal y conociendo, por consiguiente, las

transformaciones elementales que reducen $\lambda E - A$ a $\lambda E - B$, podemos encontrar, de acuerdo con el p. 13.4, unas λ-matrices $U(\lambda)$ y $V(\lambda)$ tales que

$$\lambda E - A = U(\lambda)(\lambda E - B)V(\lambda).$$

Calculando el resto a la derecha R de la división de $V(\lambda)$ por $\lambda E - A$, tendremos, según (9), $A = R^{-1}BR$, es decir, R será una de las matrices transformadoras que buscamos.

Notemos que para determinar R no es necesario realizar de hecho la división de $V(\lambda)$ por $\lambda E - A$. En efecto, representando $V(\lambda)$ en la forma

$$V(\lambda) = V_0\lambda^k + V_1\lambda^{k-1} + \ldots + V_k$$

y aplicando el esquema de división del p. 15.1, obtenemos

$$R = V_0A^k + V_1A^{k-1} + \ldots + V_k. \tag{10}$$

Análogamente, representando el polinomio $U(\lambda)$ en la forma «izquierda»

$$U(\lambda) = \lambda^l U_0 + \lambda^{l-1}U_1 + \ldots + U_l$$

y realizando la división a la izquierda por $\lambda E - A$, obtenemos para el resto a la izquierda S la expresión

$$S = A^lU_0 + A^{l-1}U_1 + \ldots + U_l. \tag{11}$$

Estas fórmulas para los restos de la división de un λ-polinomio matricial por el binomio $\lambda E - A$ son totalmente análogas a la fórmula de Bezout $r = f(a)$ para el resto de la división de un polinomio corriente $f(\lambda)$ por el binomio $\lambda - a$. Por esto las fórmulas (10) y (11) a veces se denominan fórmulas matriciales de Bezout para los restos.

15.4. Forma normal de Jordan. En el p. 12.2 hemos introducido unas matrices de forma especial que hemos llamado *matrices de Jordan* y hemos estudiado algunas propiedades de las aplicaciones lineales, cuyas matrices en un sistema de coordenadas adecuado tienen la forma de Jordan. Sin embargo, hemos dejado sin resolver el problema principal acerca de las condiciones en las que la matriz de una aplicación puede ser reducida a la forma de Jordan. Pero ahora tenemos ya todos los medios necesarios para resolver este problema.

TEOREMA 4. *Toda matriz cuadrada sobre el cuerpo de los números complejos, así como sobre cualquier otro cuerpo conmutativo algebraicamente cerrado, es semejante a una matriz de la forma de Jordan. Dos matrices de Jordan son semejantes si, y sólo si, están compuestas por las mismas células de Jordan y difieren una de la otra a lo sumo en la disposición de las células a lo largo de la diagonal principal.*

Anteponemos a la demostración del teorema dos temas que tienen también interés por sí solos.

LEMA 1. *La matriz característica de una célula de Jordan tiene sólo un divisor elemental* $(\lambda-\rho)^n$, *donde n es el orden de la célula y* ρ *es su valor propio.*

La matriz característica de una célula de Jordan dada (véase el p. 12.2) es de la forma

$$\lambda E - A = \begin{bmatrix} \lambda-\rho & -1 & 0 & \dots & 0 & 0 \\ & \lambda-\rho & -1 & \dots & 0 & 0 \\ & & & \dots & & \\ & & & & \lambda-\rho & -1 \\ & & & & & \lambda-\rho \end{bmatrix}.$$

Calculemos el máximo común divisor $D_k(\lambda)$ de los menores de orden k de la matriz $\lambda E - A$. Ante todo, tenemos

$$D_n(\lambda) = |\lambda E - A| = (\lambda-\rho)^n.$$

Después, $D_{n-1}(\lambda)$ es el máximo común divisor de todos los menores de orden $n-1$. Pero entre los últimos figura el menor

$$\begin{vmatrix} -1 & 0 & \dots & 0 & 0 \\ \lambda-\rho & -1 & \dots & 0 & 0 \\ & \lambda-\rho & \dots & 0 & 0 \\ & & \dots & & \\ & & & \lambda-\rho & -1 \end{vmatrix} = (-1)^{n-1}$$

que se obtiene al suprimir la primera columna y la última fila en la matriz $\lambda E - A$. Puesto que este menor es igual a ± 1, resulta que $D_{n-1}(\lambda) = 1$. Indiquemos por $d_1(\lambda), \dots, d_n(\lambda)$ los factores invariantes de la matriz $\lambda E - A$. De las relaciones

$$D_{n-1}(\lambda) = d_1(\lambda) \dots d_{n-1}(\lambda) = 1,$$
$$D_n(\lambda) = d_1(\lambda) \dots d_{n-1}(\lambda)\, d_n(\lambda) = (\lambda-\rho)^n$$

se desprende que $d_1(\lambda) = \dots = d_{n-1}(\lambda) = 1$ y que $d_n(\lambda) = (\lambda-\rho)^n$. Por consiguiente, $\lambda E - A$ tiene sólo un divisor elemental y este divisor es igual a $(\lambda-\rho)^n$.

LEMA 2. *El sistema de divisores elementales de la matriz característica de una matriz de Jordan se compone de los divisores elementales de sus células de Jordan y determina unívocamente, salvo el orden de secuencia de las células a lo largo de la diagonal principal, la forma de la matriz de Jordan.*

Una matriz de Jordan es, por definición, una matriz celular diagonal con células de Jordan a lo largo de la diagonal principal. Por ello, la matriz característica de una matriz de Jordan se descompone en las matrices características de las células de Jordan aisladas. De aquí se deduce, en virtud del p. 14.2, que el sistema de divisores elementales de la matriz característica de una matriz

de Jordan consta de los divisores elementales de las matrices características de cada una de las células de Jordan, habiendo un divisor por cada una de estas células. Luego, el sistema de divisores elementales de la matriz característica de una matriz de Jordan determina la forma de esta matriz unívocamente, salvo el orden de secuencia de las células a lo largo de la diagonal principal.

Las matrices características de matrices semejantes son equivalentes y, por lo tanto, tienen los mismos sistemas de divisores elementales. De aquí se deduce que las matrices semejantes de Jordan deben estar formadas por células iguales de Jordan y, para concluir la demostración del teorema 4, resta sólo saber cómo obtener a partir de toda matriz dada A la matriz de Jordan semejante a ella. Sea $(\lambda-\rho_1)^{n_1}, \ldots, (\lambda-\rho_s)^{n_s}$ el conjunto completo de los divisores elementales de la matriz característica $\lambda E-A$. Indiquemos por B la matriz celular diagonal, cuyas células diagonales son las células de Jordan con los divisores elementales señalados. Luego, la matriz $\lambda E-B$ tendrá los mismos divisores elementales que tiene $\lambda E-A$. Pero entonces, según el p. 14.1, las matrices $\lambda E-A$ y $\lambda E-B$ son equivalentes y de aquí se desprende, en virtud del p. 15.3, que la matriz A es semejante a la matriz de Jordan B. Hemos demostrado el teorema.

Los razonamientos expuestos ofrecen también una respuesta a la pregunta de cómo hallar a partir de una matriz dada A su matriz semejante de Jordan. Para ello es suficiente formar la matriz característica $\lambda E-A$, reducirla mediante transformaciones elementales a la forma canónica diagonal, descomponer en factores los polinomios diagonales, hallar los divisores elementales y construir a partir de los mismos la matriz de Jordan. Sea, por ejemplo,

$$A=\begin{bmatrix} 3 & 1 & -3 \\ -7 & -2 & 9 \\ -2 & -1 & 4 \end{bmatrix}.$$

Formamos la matriz característica

$$\lambda E-A=\begin{bmatrix} \lambda-3 & -1 & 3 \\ 7 & \lambda+2 & -9 \\ 2 & 1 & \lambda-4 \end{bmatrix}$$

y determinamos sus factores invariantes. Es fácil ver que estos factores son 1, 1, $(\lambda-1)(\lambda-2)^2$. Por consiguiente, los divisores elementales serán $\lambda-1$ y $(\lambda-2)^2$ y la matriz de Jordan es

$$B=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}.$$

Para terminar hagamos una observación más. Si los divisores elementales de la matriz $\lambda E - A$ resultan de primer grado, las células de Jordan de la matriz de Jordan correspondiente B serán de primer orden, es decir, la matriz B será diagonal. Recíprocamente, si la correspondiente matriz de Jordan es diagonal, los divisores elementales serán de primer grado. Por consiguiente, *para que una matriz dada sea semejante a una diagonal es necesario y suficiente que todos los divisores elementales de su matriz característica sean de primer grado.*

15.5. Forma normal natural. En el cuerpo de los números complejos toda matriz es equivalente a una matriz de Jordan determinada. Si el campo principal K es un cuerpo conmutativo arbitrario, la reducción a la forma de Jordan puede resultar imposible. En todo caso, para que esta reducción sea posible es necesario, y como puede verse fácilmente también es suficiente, que el polinomio característico de la matriz se descomponga sobre K en factores lineales. Es por esto que surge el problema de indicar una forma normal a la que puede ser reducida la matriz sobre el cuerpo conmutativo del que se toman sus elementos. Existe una cantidad infinita de formas normales de esta índole. Entre éstas se obtiene con mayor facilidad la forma normal llamada *natural*.

Sea

$$f(\lambda) = \alpha_0 + \alpha_1\lambda + \ldots + \alpha_{n-1}\lambda^{n-1} + \lambda^n$$

un polinomio cualquiera de grado no nulo y de coeficiente principal igual a la unidad. Aceptaremos que los coeficientes del polinomio $f(\lambda)$ pertenecen a un cuerpo conmutativo K. La matriz

$$A = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & & \ldots & 0 & 1 \\ -\alpha_0 & -\alpha_1 & -\alpha_2 & \ldots & -\alpha_{n-2} & -\alpha_{n-1} \end{bmatrix}$$

se llama matriz *asociada* del polinomio $f(\lambda)$.

LEMA 3. *Si $f(\lambda)$ es un polinomio de grado no nulo y de coeficiente principal 1 y si A es su matriz asociada, los factores invariantes de la matriz característica $\lambda E - A$ son iguales a 1, 1, ..., 1, $f(\lambda)$.*

En efecto, si $f(\lambda) = \alpha_0 + \alpha_1\lambda + \ldots + \alpha_{n-1}\lambda^{n-1} + \lambda^n$, se tiene

$$\lambda E - A = \begin{bmatrix} \lambda & -1 & 0 & \ldots & 0 & 0 \\ 0 & \lambda & -1 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & \lambda & -1 \\ \alpha_0 & \alpha_1 & \alpha_2 & \ldots & \alpha_{n-2} & \lambda + \alpha_{n-1} \end{bmatrix}.$$

Suprimiendo la primera columna y la última fila obtenemos un menor igual a $(-1)^{n-1}$. Por consiguiente, el máximo común divisor $D_{n-1}(\lambda)$ de los menores de orden $n-1$ es igual a la unidad. En cuanto a $D_n(\lambda)$, este polinomio es igual al determinante de la matriz $\lambda E - A$. Desarrollando este determinante según los elementos de la última fila, encontraremos directamente que

$$D_n(\lambda) = \alpha_0 + \alpha_1\lambda + \ldots + \alpha_{n-1}\lambda^{n-1} + \lambda^n.$$

De $D_{n-1}(\lambda) = 1$ se deduce que $d_1(\lambda) = \ldots = d_{n-1}(\lambda) = 1$ y de $D_n(\lambda) = f(\lambda)$ se desprende que $d_n(\lambda) = f(\lambda)$ que es lo que se quería demostrar.

Se dice que la matriz A es *de forma normal natural*, si A se descompone en células $A_1, A_2, \ldots, A_s$ que son las matrices asociadas de unos polinomios

$f_1(\lambda)$, $f_2(\lambda)$, ..., $f_s(\lambda)$ cada uno de los cuales es divisible por el anterior. Probemos que los factores invariantes de la matriz A son iguales a 1, ..., 1, $f_1(\lambda)$, $f_2(\lambda)$, ..., $f_s(\lambda)$, donde el número de unidades es igual a la suma de los grados de los polinomios $f_1(\lambda)$, ..., $f_s(\lambda)$ disminuida en s. Efectivamente, la matriz característica es de la forma

$$\lambda E - A = \begin{bmatrix} \lambda E_1 - A_1 & & & \\ & \lambda E_2 - A_2 & & \\ & & \ddots & \\ & & & \lambda E_s - A_s \end{bmatrix}.$$

Según el lema, toda célula $\lambda E_i - A_i$ puede ser reducida mediante transformaciones elementales a la forma diagonal en la que los elementos diagonales serán 1, ..., 1, $f_i(\lambda)$. Cambiando después en la matriz $\lambda E - A$ el orden de las filas y de las columnas, la reduciremos a la forma diagonal con los elementos diagonales 1, ..., 1 $f_1(\lambda)$, ..., $f_s(\lambda)$. Puesto que todo elemento aquí es divisible por el anterior, ésta será la forma canónica diagonal y los elementos 1, ..., 1, $f_1(\lambda)$, ..., $f_s(\lambda)$ serán los factores invariantes de la matriz $\lambda E - A$. Como quiera que el orden de la matriz A era igual a la suma de los grados de los polinomios $f_1(\lambda)$, ..., $f_s(\lambda)$ y el número total de factores invariantes es igual al orden de la matriz, resulta que el número de unidades que figuran entre los factores invariantes coincide con el señalado.

De nuestro razonamiento se desprende, en particular, que la forma normal natural queda unívocamente determinada por los factores invariantes de su matriz característica.

Ahora podemos demostrar en unas palabras el teorema siguiente.

TEOREMA 5. *Toda matriz A con elementos de un cuerpo conmutativo K se reduce sobre este cuerpo a una forma normal natural, y sólo una.*

Sean 1, ..., 1, $f_1(\lambda)$, ..., $f_s(\lambda)$ los factores invariantes de la matriz $\lambda E - A$. Puesto que la suma de los grados de todos los factores invariantes debe ser igual al orden de la matriz A, resulta que el número de unidades que figuran aquí es igual a la suma de los grados de los polinomios $f_1(\lambda)$, ..., $f_s(\lambda)$ disminuida en s. Construyamos para cada uno de los polinomios $f_i(\lambda)$ su matriz asociada B_i y consideremos la matriz celular diagonal de células diagonales B_1, ..., B_s. Como todo polinomio $f_i(\lambda)$ es divisible por el anterior, B es la matriz normal natural. Según hemos demostrado anteriormente, los factores invariantes de la matriz $\lambda E - B$ son iguales a 1, ..., 1, $f_1(\lambda)$, ..., $f_s(\lambda)$ y coinciden, por consiguiente, con los factores invariantes de la matriz $\lambda E - A$. De aquí se deduce, en virtud del teorema 3, que A es semejante a B. Con esto queda demostrada la posibilidad de reducir la matriz A a la forma normal natural. La unicidad resulta de que la matriz natural B se determina unívocamente por los factores invariantes, que, a su vez, se determinan unívocamente por la matriz A.

15.6. Otras formas normales. La ventaja de la forma normal natural estriba en que es absolutamente unívoca: tanto las propias células diagonales como el orden en que están dispuestas a lo largo de la diagonal principal se determinan unívocamente. Entre los defectos figura el que esta forma no ofrece la reducción a las células de menor orden posible así como el que no comprende la forma de Jordan como un caso particular.

El primero de estos defectos puede ser superado del modo siguiente. Convengamos en llamar una matriz B casinatural, si B se descompone en células B_1, ..., B_s que representan las matrices asociadas de polinomios de tipo $[\varepsilon_1(\lambda)]^{m_1}$, ..., $[\varepsilon_s(\lambda)]^{m_s}$, donde $\varepsilon_1(\lambda)$, ..., $\varepsilon_s(\lambda)$ son polinomios irreducibles sobre el campo principal K de coeficientes principales iguales a 1. Puesto que B_i es la matriz asociada del polinomio $[\varepsilon_i(\lambda)]^{m_i}$, los factores invariantes de la

matriz característica $\lambda E_i - B_i$ son iguales a 1, ..., 1, $[\varepsilon_i(\lambda)]^{m_i}$. Por consiguiente, $\lambda E_i - B_i$ tiene un único divisor elemental $[\varepsilon_i(\lambda)]^{m_i}$ y los divisores elementales de la matriz $\lambda E - B$ serán $[\varepsilon_1(\lambda)]^{m_1}$, ..., $[\varepsilon_s(\lambda)]^{m_s}$. Puesto que el rango de la matriz $\lambda E - B$ es igual a su orden y es igual a la suma de los grados de los polinomios $[\varepsilon_i(\lambda)]^{m_i}$, resulta que la matriz casinatural se determina por sus divisores elementales univocamente, salvo el orden de disposición de las células a lo largo de la diagonal principal.

Igual que en el punto anterior de aquí se desprende directamente que toda matriz A formada por elementos de un cuerpo conmutativo K se reduce sobre este cuerpo a la forma normal casinatural. Esta forma queda determinada por la matriz A univocamente, salvo el orden en el que siguen las células a lo largo de la diagonal principal.

Las células de la forma casinatural no pueden ser descompuestas sobre el cuerpo conmutativo K en células de menor orden, ya que si esta descomposición fuese posible, resultaría que la matriz característica de la célula tendría por lo menos dos divisores elementales en lugar de uno.

Para concluir examinemos la forma normal que puede ser considerada como una generalización de la forma de Jordan al caso de subcuerpos conmutativos arbitrarios K del cuerpo de los números complejos [1]. Sea $\varepsilon(\lambda)$ un polinomio irreducible con coeficientes de K. Aceptaremos que el polinomio $\varepsilon(\lambda)$ es de grado no nulo y que su coeficiente principal es igual a la unidad.

LEMA 4. *Si el polinomio característico de una matriz A formada por elementos de un cuerpo conmutativo numérico K es irreducible sobre K y es igual a $\varepsilon(\lambda)$, la matriz celular*

$$B = \begin{bmatrix} A & E & 0 & \dots & 0 \\ & A & E & \dots & 0 \\ & & \cdot & \cdot & \cdot \\ & & & A & E \\ & & & & A \end{bmatrix},$$

donde E es la matriz unidad, tiene una matriz característica que posee un único divisor elemental $[\varepsilon(\lambda)]^m$, donde m es el número de células diagonales de la matriz B.

Sea $\lambda E - A = P$; entonces

$$\lambda E - B = \begin{bmatrix} P & -E & 0 & \dots & 0 & 0 \\ & P & -E & \dots & 0 & 0 \\ & & & \cdot & \cdot & \cdot \\ & & & & P & -E \\ & & & & & P \end{bmatrix}. \tag{12}$$

Si agregamos ahora a las células de una columna cualquiera de la matriz (12) las células de otra cualquiera de sus columnas multiplicadas por una λ-matriz arbitraria S, esta operación equivaldrá, obviamente, a una serie de transformaciones elementales de tipo II o IV realizadas con la matriz $\lambda E - B$, de modo que el resultado será una matriz equivalente a $\lambda E - B$. Tengamos en cuenta esta observación y realicemos sucesivamente las siguientes transformaciones con la matriz $\lambda E - B$: agreguemos a su primera columna la segunda columna multiplicada por P; agreguemos a la primera columna de la matriz nueva su tercera columna multiplicada por P^2, a su segunda columna la tercera multiplicada por P, etc. Después de estas transformaciones obtendremos la

[1]) Hablando más generalmente, al caso de cualesquiera cuerpos conmutativos perfectos. Pero entonces, en lugar del cuerpo de los números complejos habrá que tomar la adherencia algebraica del cuerpo conmutativo correspondiente.

matriz

$$\begin{bmatrix} 0 & -E & 0 & \dots & 0 \\ 0 & 0 & -E & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -E \\ P^m & P^{m-1} & P^{m-2} & \dots & P \end{bmatrix}.$$

Agregando ahora a la última fila la primera, la segunda, ..., filas multiplicadas, respectivamente, por P^{m-1}, P^{m-2}, ... y cambiando el orden de las columnas, obtendremos la matriz

$$\begin{bmatrix} -E & & & \\ & \cdot & & \\ & & \cdot & \\ & & & -E_{P^m} \end{bmatrix}. \tag{13}$$

Puesto que la matriz (13) tiene la forma celular diagonal y sus células iniciales coinciden, salvo el signo, con matrices unidades, resulta que los divisores elementales de la matriz (13) coinciden con los divisores elementales de la matriz $P^m = (\lambda E - A)^m$ que pasamos ahora a examinar.

Por hipótesis, el polinomio característico de la matriz A es igual a $\varepsilon(\lambda)$ y es irreducible sobre el cuerpo conmutativo K. De aquí se desprende que todas sus raíces en el cuerpo de los números complejos son diferentes y que, por consiguiente, existe una matriz compleja T tal que TAT^{-1} es de forma diagonal y que

$$\lambda E - TAT^{-1} = \begin{bmatrix} \lambda-\alpha_1 & & & \\ & \lambda-\alpha_2 & & \\ & & \cdot & \\ & & & \lambda-\alpha_n \end{bmatrix}.$$

La matriz P^m es equivalente a la matriz $TP^mT^{-1} = (TPT^{-1})^m$ y la última es de la forma

$$[T(\lambda E - A)T^{-1}]^m = (\lambda E - TAT^{-1})^m = \begin{bmatrix} (\lambda-\alpha_1)^m & & & \\ & \cdot & & \\ & & \cdot & \\ & & & (\lambda-\alpha_n)^m \end{bmatrix}.$$

Por esto, los divisores elementales de la matriz P^m en el cuerpo de los números complejos son iguales a $(\lambda-\alpha_1)^m$, ..., $(\lambda-\alpha_n)^m$. Todos ellos pertenecen a diferentes polinomios irreducibles. Los factores invariantes de la matriz P^m son $1, \dots, 1, (\lambda-\alpha_1)^m \dots (\lambda-\alpha_n)^m = [\varepsilon(\lambda)]^m$. Pero el polinomio $\varepsilon(\lambda)$ es irreducible sobre el cuerpo conmutativo K y, por consiguiente, la matriz $\lambda E - P^m$ tiene en este cuerpo conmutativo sólo un divisor elemental, a saber, $[\varepsilon(\lambda)]^m$. Es decir, la matriz (13) y con ella también la matriz (12) tienen sólo un divisor elemental $[\varepsilon(\lambda)]^m$. Hemos demostrado el lema.

Si A es la matriz asociada de $\varepsilon(\lambda)$, diremos que la matriz B es *la célula generalizada de Jordan* correspondiente al divisor elemental $[\varepsilon(\lambda)]^m$. Asimismo diremos que una matriz tiene *la forma generalizada de Jordan*, si se descompone en células generalizadas de Jordan. Está claro que una matriz generalizada de Jordan se determina plenamente por sus divisores elementales.

TEOREMA 6. *Toda matriz A formada por elementos de un cuerpo conmutativo K se reduce sobre este cuerpo a la forma generalizada de Jordan. Está forma*

queda determinada por la matriz A unívocamente, salvo el orden de disposición de las células a lo largo de la diagonal principal.

En efecto, sean $[\varepsilon_1(\lambda)]^{m_1}, \ldots, [\varepsilon_s(\lambda)]^{m_s}$ los divisores elementales de la matriz $\lambda E - A$ en el cuerpo conmutativo K. Construyamos para todo polinomio $[\varepsilon_i(\lambda)]^{m_i}$ la correspondiente célula generalizada de Jordan B_i y consideremos la matriz celular diagonal B con las células B_i a lo largo de la diagonal principal.

En virtud del lema 4, los divisores elementales de la matriz $\lambda E - B$ son iguales a los respectivos divisores elementales de la matriz $\lambda E - A$. Puesto que $\lambda E - B$ y $\lambda E - A$ son, además, del mismo rango y del mismo orden, resulta que A es semejante a B. La unicidad se deduce de que los divisores elementales de la matriz $\lambda E - A$ determinan unívocamente la matriz B.

Si el campo principal K es cuerpo de los números complejos, todos los polinomios irreducibles son del primer grado. Por consiguiente, las matrices A y B de las células generalizadas de Jordan serán del primer orden, de modo que las células generalizadas se convierten en las células corrientes de Jordan.

Consideremos también el caso en que el campo principal K es el cuerpo de todos los números reales. Los polinomios irreducibles sobre K serán de dos tipos: 1) polinomios de primer grado $\lambda - \rho$; las respectivas células generalizadas de Jordan serán células corrientes de Jordan y 2) $\varepsilon(\lambda) = \lambda^2 + p\lambda + q$, donde $p^2 - 4q < 0$; la célula asociada A es de la forma

$$\begin{bmatrix} 0 & 1 \\ -q & -p \end{bmatrix} \tag{14}$$

y la célula generalizada de Jordan correspondiente a $(\lambda^2 + p\lambda + q)^m$ tiene la forma de la matriz B del lema 4, donde en lugar de A se puede tomar una matriz de otro tipo

$$A = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix};$$

α y β son los coeficientes de la parte real y de la parte imaginaria, respectivamente, de la raíz compleja del polinomio $\lambda^2 + p\lambda + q$. Esta sustitución es posible, ya que el polinomio característico de la matriz

$$\lambda E - A = \begin{bmatrix} \lambda - \alpha & -\beta \\ \beta & \lambda - \alpha \end{bmatrix}$$

es igual a $\lambda^2 + p\lambda + q$.

Ejemplos y problemas

1. Determínese cuáles de las matrices

$$\begin{bmatrix} 3 & 1 & -3 \\ -4 & -2 & 6 \\ -1 & -1 & 5 \end{bmatrix}, \quad \begin{bmatrix} 42 & 130 & 25 \\ -8 & -24 & -5 \\ -23 & -73 & -13 \end{bmatrix}, \quad \begin{bmatrix} 2 & 1 & -1 \\ 7 & 4 & -22 \\ 2 & 1 & -5 \end{bmatrix},$$

$$\begin{bmatrix} 8 & -13 & 16 \\ -8 & 18 & -22 \\ -11 & 22 & -27 \end{bmatrix}, \quad \begin{bmatrix} 20 & -89 & -32 \\ 11 & -51 & 20 \\ 20 & -95 & -36 \end{bmatrix} \text{ y } \begin{bmatrix} 2 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & -1 & 3 \end{bmatrix}$$

son semejantes. Hállense sus matrices de Jordan semejantes.

2. Demuéstrese que el polinomio característico de una matriz A es igual al producto de todos los factores invariantes y que el polinomio mínimo es igual al último de los factores invariantes de la matriz característica $\lambda E - A$.

3. Hállense los polinomios mínimos de las matrices indicadas en el problema 1.

4. Demuéstrese que sobre el cuerpo de los números complejos toda matriz cuadrada es semejante a una matriz diagonal si, y sólo si, su polinomio mínimo no tiene raíces múltiples.

5. Demuéstrese que las matrices recíprocamente transpuestas A y A' son siempre semejantes.

6. Si A es una matriz regular y B es una matriz cualquiera, la matriz AB es semejante a la matriz BA. ¿Serán semejantes AB y BA siendo arbitrarias las matrices A y B?

7. Hállense las formas normales de las matrices

$$\begin{bmatrix} -3 & -1 & -3 \\ 22 & 9 & -27 \\ 5 & 2 & -6 \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} 1 & -1 & 1 & 3 \\ -1 & 1 & 3 & -1 \\ -1 & -3 & -1 & -1 \\ 3 & 1 & -1 & -1 \end{bmatrix}$$

sobre el cuerpo de los números racionales, sobre el cuerpo de los números reales y sobre el cuerpo de los números complejos.

8. Muéstrese que para obtener la forma normal sobre el cuerpo de los números complejos se puede tomar en lugar de las células de Jordan células de tipo

$$\begin{bmatrix} \rho & \alpha & 0 & \dots & 0 & 0 \\ & \rho & \alpha & \dots & 0 & 0 \\ & & \rho & \dots & 0 & 0 \\ & & & \dots & & \\ & & & & \rho & \alpha \\ & & & & & \rho \end{bmatrix},$$

donde α es un número fijo cualquiera diferente de cero.

9. CARACTERÍSTICA DE SEGRE. Sea $\mathcal{A}$ una aplicación lineal de un espacio vectorial complejo y sean $\rho_1, \dots, \rho_s$ distintos valores propios de esta aplicación. Indiquemos por $\sigma_{1k}, \sigma_{2k}, \dots, \sigma_{jk}$ los órdenes de las células de Jordan en la forma normal de Jordan de $\mathcal{A}$, correspondientes al valor propio ρ_k. El símbolo

$$[(\sigma_{11}, \sigma_{21}, \dots)(\sigma_{12}, \sigma_{22}, \dots) \dots (\sigma_{1s}, \sigma_{2s}, \dots)]$$

se denomina *característica de Segre de la aplicación* $\mathcal{A}$. Por ejemplo, las matrices

$$\begin{bmatrix} 1 & 1 & & \\ 0 & 1 & & \\ & & 1 & \\ & & & 2 \end{bmatrix}, \quad \begin{bmatrix} 2 & 1 & & \\ 0 & 2 & & \\ & & 2 & 1 \\ & & 0 & 2 \end{bmatrix}, \quad \begin{bmatrix} 2 & 1 & 0 & \\ & 2 & 1 & \\ & & 2 & \\ & & & 3 \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} 2 & 1 & 0 & 0 \\ & 2 & 1 & 0 \\ & & 2 & 1 \\ & & & 2 \end{bmatrix}$$

tienen, respectivamente, las siguientes características de Segre: $[(2,1)\ (1)]$, $[(2,2)]$, $[(3)\ (1)]$ y $[(4)]$. Calcúlense las características de Segre de las matrices indicadas en el problema 1.

10. CARACTERÍSTICA DE WEYR. Sea $\mathcal{A}$ la misma aplicación del problema anterior. Indiquemos por

$$\alpha_{i1}, \alpha_{i1}+\alpha_{i2}, \dots, \alpha_{i1}+\alpha_{i2}+\dots+\alpha_{ip}$$

los defectos de las aplicaciones $\mathcal{A}-\rho_i\mathcal{E}$, $(\mathcal{A}-\rho_i\mathcal{E})^2, \dots, (\mathcal{A}-\rho_i\mathcal{E})^p$, donde ρ es la multiplicidad del valor propio ρ_i. La fila $\{\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{ip}\}$ se denomina *característica de Weyr de la aplicación* $\mathcal{A}$ correspondiente al valor propio ρ_i. Demuéstrese que las características de Weyr de las aplicaciones indicadas en el problema 9 son iguales, respectivamente, a

$$[\{2, 1\}, \{1\}], [\{2, 2\}], [\{1, 1, 1\}, \{1\}] \text{ y } [\{1, 1, 1, 1\}].$$

11. Demuéstrese que la característica de Segre y la característica de Weyr correspondientes a un mismo valor propio están relacionadas del modo siguiente.

Supongamos que la característica de Segre es (6, 4, 3, 3). Consideremos el diagrama de puntos

El número de puntos en las sucesivas columnas de este diagrama será precisamente la característica de Weyr. Para el caso considerado es igual a {4, 4, 4, 2, 1, 1}.

12. Toda matriz con divisores elementales irreducibles se denomina *semisimple*. Demuéstrese que cualquier matriz A puede ser descompuesta en la suma de una matriz semisimple y otra nilpotente permutables entre sí y que esta descomposición es única.

§ 16. Funciones de matrices

En este parágrafo serán considerados aquellos problemas del cálculo de matrices en la solución de los cuales se emplea la posibilidad de reducir las matrices a la forma normal de Jordan. De acuerdo con esto se aceptará que el campo principal es el cuerpo de todos los números complejos.

16.1. Polinomio en una matriz de Jordan. Las funciones de matrices más sencillas son los polinomios. Más tarde daremos la definición general de función de una matriz, mientras que ahora daremos la expresión explícita de un polinomio en una matriz que tiene la forma normal de Jordan. Consideremos primero una célula aislada de Jordan de orden n

$$A=\begin{bmatrix} \rho & 1 & 0 & \dots & 0 & 0 \\ & \rho & 1 & \dots & 0 & 0 \\ & & & \dots\dots & & \\ & & & & \rho & 1 \\ & & & & & \rho \end{bmatrix}. \tag{1}$$

Demostremos que para cualesquiera valores naturales de m es válida la fórmula

$$A^m=\begin{bmatrix} \rho^m & \binom{m}{1}\rho^{m-1} & \dots & \binom{m}{n-1}\rho^{m-n+1} \\ & \rho^m & \dots & \binom{m}{n-2}\rho^{m-n+2} \\ & & \dots\dots\dots\dots & \\ & & & \rho^m \end{bmatrix}, \tag{2}$$

donde se ha tomado

$$\binom{m}{k}=\frac{m(m-1)\dots(m-k+1)}{1\cdot 2\dots k}.$$

El método de demostración más sencillo es por inducción según m. Para $m=1$ la fórmula (2) coincide con (1) y por lo tanto

es verídica. Por otro lado, si la igualdad (2) es válida para un valor de m, multiplicándola por A, obtendremos, mediante el cálculo directo, que la fórmula (2) es también válida para A^{m+1}.

Sea ahora $f(\lambda)$ un polinomio en λ:

$$f(\lambda) = \alpha_0 + \alpha_1\lambda + \alpha_2\lambda^2 + \ldots + \alpha_k\lambda^k.$$

Tenemos, por definición,

$$f(A) = \alpha_0 E + \alpha_1 A + \alpha_2 A^2 + \ldots + \alpha_k A^k.$$

Introduciendo aquí en lugar de las matrices A^m sus valores de (2), veremos que en la i-ésima fila y en la $(i+s)$-ésima columna de la matriz $f(A)$ aparece la expresión

$$\sum_{m=0}^{k} \alpha_m \frac{m(m-1)\ldots(m-s)}{1\cdot 2\ldots s}\rho^{m-s} = \frac{1}{1\cdot 2\ldots s} f^{(s)}(\rho).$$

Por consiguiente, obtenemos definitivamente que

$$f(A) = \begin{bmatrix} f(\rho) & \frac{1}{1!}f'(\rho) & \frac{1}{2!}f''(\rho) & \ldots & \frac{1}{(n-1)!}f^{(n-1)}(\rho) \\ & f(\rho) & \frac{1}{1!}f'(\rho) & \ldots & \frac{1}{(n-2)!}f^{(n-2)}(\rho) \\ & & & \ldots\ldots\ldots & \\ & & & & f(\rho) \end{bmatrix}. \qquad (3)$$

Hemos calculado el valor de un polinomio en una célula de Jordan. Pero una matriz general de Jordan A es una suma directa de células aisladas de Jordan:

$$A = A_1 \dotplus A_2 \dotplus \ldots \dotplus A_s$$

y según el p. 1.4 tenemos

$$f(A) = f(A_1) \dotplus f(A_2) \dotplus \ldots \dotplus f(A_s). \qquad (4)$$

Aquí $f(A_1), \ldots f(A_s)$ son polinomios en células aisladas de Jordan y sus expresiones vienen dadas por la fórmula (3). Este resultado se puede aplicar también para el cálculo de polinomios en matrices A que no tienen la forma de Jordan. En efecto, determinamos primero una matriz T tal que la matriz $T^{-1}AT = B$ tenga la forma normal de Jordan, calculamos después $f(B)$ empleando las fórmulas (3) y (4) y, finalmente, teniendo en cuenta la relación

$$f(A) = f(TBT^{-1}) = Tf(B)T^{-1}$$

(p. 3.1) obtenemos el valor de $f(A)$.

16.2. Funciones escalares. El concepto general de funciones matriciales se define de la misma forma absolutamente que el concepto de funciones numéricas corrientes. A saber, consideremos un conjunto

de matrices $\mathfrak{M}$. Si a toda matriz A de $\mathfrak{M}$ se pone en correspondencia una matriz B, se dice que B es una función de A definida sobre $\mathfrak{M}$. Queremos ahora poner en correspondencia a toda función numérica corriente $\rho = f(\lambda)$ —que está definida sobre un conjunto de números complejos y que satisface las condiciones indicadas más abajo— una determinada función matricial $f(A)$. Esta correspondencia se obtiene del modo siguiente. Sean dadas una función numérica $\rho = f(\lambda)$ y una matriz arbitraria A. Indiquemos por $\rho_1, \rho_2, \ldots, \rho_s$ los diferentes valores propios de la matriz A. Reduzcamos A a la forma normal de Jordan

$$T^{-1}AT = B = B_1 + B_2 + \ldots + B_t,$$

donde $B_1, \ldots, B_t$ son células de Jordan, y consideremos una de estas células, por ejemplo, la célula

$$B_i = \begin{bmatrix} \rho_i & 1 & 0 \ldots 0 \\ & \rho_i & 1 \ldots 0 \\ & & \ldots \\ & & \rho_i \end{bmatrix} \tag{5}$$

que corresponde al divisor elemental $(\lambda - \rho_i)^{n_i}$. Si la función $f(\lambda)$ está definida en una vecindad del punto ρ_i y tiene derivadas finitas $f'(\rho_i), \ldots, f^{(n_i-1)}(\rho_i)$, tomamos por definición

$$f(B_i) = \begin{bmatrix} f(\rho_i) & \frac{1}{1!}f'(\rho_i) & \ldots & \frac{1}{(n_i-1)!}f^{(n_i-1)}(\rho_i) \\ & f(\rho_i) & \ldots & \frac{1}{(n_i-2)!}f^{(n_i-2)}(\rho_i) \\ & & \ldots & \ldots \\ & & & f(\rho_i) \end{bmatrix}. \tag{6}$$

Además, si $f(\lambda)$ está definida en una vecindad de cada uno de los puntos $\rho_1, \ldots, \rho_s$ y tiene en estas vecindades derivadas de orden adecuado, tomamos también

$$f(B) = f(B_1) + f(B_2) + \ldots + f(B_t) \tag{7}$$

y

$$f(A) = Tf(B)T^{-1} = T(f(B_1) + \ldots + f(B_t))T^{-1}. \tag{8}$$

La matriz $f(A)$ se llama *valor* de la función $f(\lambda)$ para $\lambda = A$. Más abajo se demuestra que $f(A)$ no depende de cómo se reduce la mátriz A a la forma normal[1] y es, por consiguiente, una función matricial de A. Esta función se denomina *correspondiente* a la función numérica $f(\lambda)$. Está claro que no todas, ni mucho menos, funciones matriciales poseen las correspondientes funciones numéricas.

[1] Es decir, de cómo se escoge la matriz T.

Aquellas para las cuales existen funciones numéricas correspondientes se denominan *funciones escalares*.

Indiquemos algunas propiedades elementales de las funciones escalares:

A) Si $f(\lambda)$ es un polinomio en λ, el valor de la función escalar $f(A)$ coincide con el valor del polinomio $f(\lambda)$ para $\lambda = A$ en el sentido del p.1.2.

Efectivamente, la propia definición de las funciones escalares se ha realizado de modo que, para el caso de polinomios, coincida con la antigua.

B) Sea A una matriz y sean $f_1(\lambda)$ y $f_2(\lambda)$ unas funciones numéricas para las cuales tienen sentido las expresiones $f_1(A)$ y $f_2(A)$. Si $f(\lambda) = f_1(\lambda) + f_2(\lambda)$, también $f(A)$ tiene sentido y $f(A) = f_1(A) + f(A_2)$.

C) Si A es una matriz, $f_1(\lambda)$ y $f_2(\lambda)$ son unas funciones numéricas tales que $f_1(A)$ y $f_2(A)$ tienen sentido y si $f(\lambda) = f_1(\lambda) f_2(\lambda)$, también $f(A)$ tiene sentido y $f(A) = f_1(A) f_2(A)$.

Las demostraciones de las propiedades B) y C) son semejantes y por ello nos limitaremos al caso de la propiedad C). Para calcular los valores $f_1(A)$, $f_2(A)$ y $f(A)$ debemos, por definición, reducir A a la forma normal de Jordan B y emplear las fórmulas (7) y (8). Si se logra demostrar que $f(B) = f_1(B) f_2(B)$, de la fórmula (8) resultará directamente que $f(A) = f_1(A) f_2(A)$. Por otro lado, se tiene

$$f(B) = f(B_1) + f(B_2) + \ldots + f(B_t),$$

$$f_1(B) f_2(B) = f_1(B_1) f_2(B_1) + \ldots + f_1(B_t) f_2(B_t)$$

y, por consiguiente, todo se reduce a la demostración de las igualdades

$$f(B_i) = f_1(B_i) f_2(B_i) \quad (i = 1, 2, \ldots, t),$$

donde B_i son células de Jordan. Tomando los valores de $f_1(B_i)$ y $f_2(B_i)$ según las fórmulas (6) y multiplicándolos, veremos que en la k-ésima fila y en la $(k+l)$-ésima columna de la matriz $f_1(B_i) f_2(B_i)$ aparece el elemento igual a

$$f_1(\rho) \cdot \frac{1}{l!} f_2^{(l)}(\rho) + \frac{1}{1!} f_1'(\rho) \cdot \frac{1}{(l-1)!} f_2^{(l-1)}(\rho) + \ldots + \frac{1}{l!} f_1^{(l)}(\rho) \cdot f_2(\rho).$$

Esta expresión puede ser representada en la forma

$$\frac{1}{l!}\left[f_1(\rho) f_2^{(l)}(\rho) + \frac{l}{1!} f_1'(\rho) f_2^{(l-1)}\rho + \ldots + f_1^{(l)}(\rho) f_2(\rho)\right]$$

que, de acuerdo con la regla de la derivada de un producto de funciones, coincide con $\frac{1}{l!} f^{(l)}(\rho)$. Por consiguiente

$$f_1(B_i) f_2(B_i) = f(B_i)$$

y la proposición C) queda demostrada.

Análogamente, empleando esta vez la regla de la derivación de una función de función, se podría demostrar que siendo $\varphi(\lambda)$ y $f(\varphi(\lambda))$ funciones numéricas que satisfacen las condiciones en las que la expresión $f(\varphi(A))$ tenga sentido y siendo $\psi(\lambda) = f(\varphi(\lambda))$ se tiene $\psi(A) = f(\varphi(A))$.

D) Sea A una matriz cuyos valores propios son $\rho_1, \rho_2, \ldots, \rho_n$ con la particularidad de que todo valor propio aparece aquí tantas veces como su multiplicidad lo indique. Si $f(\lambda)$ es una función numérica y si $f(A)$ tiene sentido, los valores propios de la matriz $f(A)$ son iguales a $f(\rho_1), f(\rho_2), \ldots, f(\rho_n)$.

En efecto, los valores propios respectivos de las matrices $f(A)$ y $T^{-1}f(A)T = f(T^{-1}AT)$ coinciden y, por lo tanto, podemos aceptar que A tiene la forma normal de Jordan. Las fórmulas (5) y (6) muestran que en este caso $f(A)$ tiene la forma triangular con la particularidad de que a lo largo de la diagonal principal de $f(A)$ figuran los números $f(\rho_1), f(\rho_2), \ldots, f(\rho_n)$. Puesto que los elementos diagonales de una matriz triangular son sus valores propios, la proposición D) queda demostrada.

Consideremos dos ejemplos. 1) Sea $f(\lambda) = \lambda^{-1}$. Esta función está definida en todo punto, salvo $\lambda = 0$, y para todos los valores de λ diferentes de cero tiene derivadas de cualquier orden. Por consiguiente, si la matriz A no posee valores propios nulos, es decir, si A es regular, resulta que $f(A)$ tiene sentido. Pero $\lambda \cdot f(\lambda) = 1$ y por lo tanto $A \cdot f(A) = E$, de donde tenemos $f(A) = A^{-1}$. Es decir, a la función λ^{-1} le corresponde la matriz inversa.

2) Sea $f(\lambda) = \sqrt{\lambda}$. Para $\lambda \neq 0$ esta función tiene derivadas finitas de cualquier orden. Por consiguiente, la expresión $\sqrt{A}$ tiene sentido para todas las matrices regulares A[1]. Tomando $\lambda = A$ en la relación

$$f(\lambda) f(\lambda) = \lambda$$

obtenemos

$$f(A) f(A) = A.$$

Hemos demostrado, por consiguiente, que de toda matriz regular se puede extraer la raíz cuadrada.

16.3. Representación de los valores de funciones por polinomios. En todos los cursos del Algebra superior se considera el problema de cómo a partir de un sistema dado de números diferentes $\rho_1, \rho_2, \ldots, \rho_s$ y de otro sistema cualquiera de números $\alpha_1, \alpha_2, \ldots, \alpha_s$ obtener un polinomio $f(\lambda)$ que en los puntos $\rho_1, \rho_2, \ldots, \rho_s$ tome

[1] Para evitar la multiformidad de $\sqrt{\lambda}$ y hacer más rigurosos los razonamientos, es suficiente efectuar en el plano complejo de la variable λ un corte desde el origen de coordenadas a lo largo de un rayo que no contenga ninguno de los valores propios de la matriz A y considerar sólo una de las ramas del radical.

los valores α_1, α_2, ..., α_s, respectivamente. La solución se ofrece en forma del conocido polinomio de interpolación de Lagrange.

En lo sucesivo será importante saber construir los polinomios cuando tanto ellos como sus derivadas hasta un orden determinado toman los valores dados en los puntos ρ_1, ρ_2, ..., ρ_s. Este problema es, por consiguiente, una generalización directa del problema anterior. Enunciemos la proposición referente a su solución en forma de un lema especial.

LEMA. *Sean ρ_1, ρ_2, ..., ρ_s unos números diferentes dados y sea dada una tabla de $(k+1)s$ números cualesquiera α_{ij}. Existe un polinomio $p(\lambda)$ que en todo punto ρ_i toma el valor α_{i0} mientras que su j-ésima derivada toma el valor α_{ij} ($i=1, 2, \ldots, s$; $j=1, \ldots, k$).*

Primero conviene construir un polinomio auxiliar $p_i(\lambda)$ que tanto él como sus derivadas hasta el orden k tomen los valores requeridos solamente en el punto ρ_i y se anulen en los demás puntos. Tomemos

$$\varphi_i(\lambda) = \beta_{i0} + \beta_{i1}(\lambda - \rho_i) + \ldots + \beta_{ik}(\lambda - \rho_i)^k,$$
$$\Phi_i(\lambda) = (\lambda - \rho_1)^{k+1} \ldots (\lambda - \rho_{i-1})^{k+1} (\lambda - \rho_{i+1})^{k+1} \ldots (\lambda - \rho_s)^{k+1},$$
$$p_i(\lambda) = \varphi_i(\lambda)\Phi_i(\lambda),$$

donde β_{i0}, β_{i1}, ..., β_{ik} son unos números por ahora indeterminados. Es evidente que para cualesquiera valores de β_{i0}, ..., β_{ik} se tiene

$$p_i(\rho_j) = p_i'(\rho_j) = \ldots = p_i^{(k)}(\rho_j) = 0 \quad (j \neq i).$$

De acuerdo a la regla de la derivación de un producto tenemos

$$p_i^{(j)}(\rho_i) = \varphi_i^{(j)}(\rho_i)\Phi_i(\rho_i) + j\varphi_i^{(j-1)}(\rho_i)\Phi_i'(\rho_i) + \ldots + \varphi_i(\rho_i)\Phi_i^{(j)}(\rho_i),$$

es decir,

$$\alpha_{ij} = j!\beta_{ij}\Phi_i(\rho_i) + j!\beta_{i,j-1}\Phi_i'(\rho_i) + \ldots + \beta_{i0}\Phi_i^{(j)}(\rho_i). \qquad (9)$$

Como $\Phi_i(\rho_i) \neq 0$, tomando $j = 0, 1, \ldots, k$ podemos determinar de las relaciones (9) sucesivamente los números β_{i0}, β_{i1}, ..., β_{ik} y, con ello, calcular $p_i(\lambda)$. El polinomio

$$p(\lambda) = p_1(\lambda) + p_2(\lambda) + \ldots + p_s(\lambda)$$

satisfará, obviamente, las condiciones del lema.

Consideremos una función numérica $f(\lambda)$ y una matriz A tal que el valor $f(A)$ esté definido. Probemos que existe entonces un polinomio $p(\lambda)$ tal que $p(A)$ es igual a $f(A)$. Indiquemos por ρ_1, ρ_2, ..., ρ_s los distintos valores propios de la matriz A. Sea n su orden. De acuerdo con el lema que acabamos de demostrar podemos construir un polinomio $p(\lambda)$ que satisfaga las condiciones siguientes[1])

$$p(\rho_i) = f(\rho_i), \quad p'(\rho_i) = f'(\rho_i), \quad \ldots, \quad p^{(n-1)}(\rho_i) = f^{(n-1)}(\rho_i) \quad (10)$$
$$(i = 1, \ldots, s).$$

[1]) Si algunas de las derivadas $f^{(j)}(\rho_i)$ sobran para la determinación de $f(A)$, los números correspondientes de (10) pueden ser sustituidos por ceros.

Para determinar el sentido de la expresión $f(A)$ sólo necesitamos conocer los valores que toman en los puntos $\rho_1, \rho_2, \ldots, \rho_s$ la función $f(\lambda)$ y sus derivadas hasta el orden $n-1$ a lo sumo. Puesto que estos valores de $f(\lambda)$ y de $p(\lambda)$ coinciden, resulta que $f(A) = p(A)$. Es decir, hemos obtenido el siguiente resultado:

TEOREMA 1. *Los valores de todas las funciones escalares de una matriz A pueden ser representados mediante polinomios en A* [1].

En particular, considerando la función $f(\lambda) = \sqrt{\lambda}$ vemos que *para toda matriz regular A existe un polinomio* $p(\lambda)$ *tal que*

$$p(A)\,p(A) = A.$$

Empleando el teorema 1 es fácil resolver el problema que hemos dejado pendiente en el punto anterior acerca de la unicidad de la determinación del valor de $f(A)$. En efecto, conociendo los valores de la función $f(\lambda)$ y de sus derivadas en los puntos $\rho_1, \ldots, \rho_s$ podemos construir el polinomio $p(\lambda)$ cuyo valor $p(A)$ no depende de cómo se reduce la matriz A a la forma normal de Jordan y coincide, al mismo tiempo, con el valor $f(A)$. Es decir, el valor $f(A)$, definido en el punto anterior mediante la reducción de la matriz A a la forma normal, no depende de cómo se realiza esta reducción.

Hagamos una observación más. Sea $f(\lambda)$ una función numérica y sea A una matriz tal que $f(A)$ tiene sentido. En virtud del teorema 1 podemos hallar un polinomio $p(\lambda)$ tal que $p(A) = f(A)$. Dada la función $f(\lambda)$, el polinomio $p(\lambda)$ depende sólo de los divisores elementales de la matriz A. Pero los divisores elementales de la matriz A y de la matriz transpuesta A' coinciden y, por ello, se tiene $p(A') = f(A')$. Es fácil deducir del punto 1.3 que $p(A') = p(A)'$. Es decir, *para toda función escalar* $f(A)$ *tenemos* $f(A') = f(A)'$

16.4. Divisores elementales de funciones. Estudiemos el problema de cómo determinar a partir de los divisores elementales de una matriz A los divisores elementales de una de sus funciones escalares $f(A)$. Reduzcamos A a la forma normal

$$T^{-1}AT = B = B_1 + B_2 + \ldots + B_t, \tag{11}$$

donde $B_1, \ldots, B_t$ son células de Jordan. Por definición,

$$f(A) = Tf(B)\,T^{-1}$$

y, por consiguiente, los divisores elementales de las matrices $f(A)$ y $f(B)$ coinciden. De (11) se desprende que

$$f(B) = f(B_1) + f(B_2) + \ldots + f(B_t);$$

[1] Notemos una vez más que, según los razonamientos expuestos en el texto, todo valor $f(A)$ de una función escalar f dada se puede representar mediante un polinomio $p(A)$. Sin embargo, para una misma función f este polinomio será diferente, según sean diferentes las matrices A.

luego, el sistema de los divisores elementales de la matriz $f(B)$ es la unión de los sistemas de los divisores elementales de las células $f(B_1), \ldots, f(B_t)$. Es decir, nuestro problema inicial se reduce al siguiente: dada una célula de Jordan B_i de divisor elemental $(\lambda-\rho_i)^{n_i}$ hallar los divisores elementales de $f(B_i)$.

En virtud de las fórmulas (5) y (6) tenemos

$$\lambda E_i - f(B_i) = \begin{bmatrix} \lambda - f(\rho_i) & -f'(\rho_i) & \ldots & -\frac{1}{(n_i-1)!} f^{(n_i-1)}(\rho_i) \\ & \lambda - f(\rho_i) & \ldots & -\frac{1}{(n_i-2)!} f^{(n_i-2)}(\rho_i) \\ & & \ldots & \ldots \\ & & & \lambda - f(\rho_i) \end{bmatrix}. \quad (12)$$

Determinemos los máximos comunes divisores $D_1(\lambda)$, $D_2(\lambda)$, ..., $D_{n_i}(\lambda)$ de los menores de primero, segundo, ..., n_i-ésimo orden de esta matriz. El mayor de ellos $D_{n_i}(\lambda)$ es igual al determinante de la matriz, es decir,

$$D_{n_i}(\lambda) = (\lambda - f(\rho_i))^{n_i}.$$

Todos los demás son divisores de $D_{n_i}(\lambda)$ y, por consiguiente, son de la forma $(\lambda-f(\rho_i))^s$. Consideremos $D_{n_i-1}(\lambda)$. Este polinomio debe ser un divisor de todos los menores de orden n_i-1 de la matriz (12) y, en particular, del menor $\Delta(\lambda)$ que se obtiene suprimiendo la primera columna y la última fila. Sin embargo, si en este menor se introduce en lugar de λ el número $f(\rho_i)$, se obtiene una matriz de forma triangular y con los elementos $-f'(\rho_i)$ a lo largo de la diagonal principal y, por consiguiente,

$$\Delta(\rho_i) = (-f'(\rho_i))^{n_i-1}. \quad (13)$$

Supongamos ahora que $f'(\rho_i) \neq 0$. La igualdad (13) muestra entonces que $\Delta(\lambda)$ no es divisible por $\lambda-f(\rho_i)$. Pero el polinomio $D_{n_i-1}(\lambda)$ debe ser un divisor común de los polinomios $\Delta(\lambda)$ y $D_{n_i}(\lambda)$, es decir, $D_{n_i-1}(\lambda) = 1$. Los demás polinomios $D_{n_i-2}(\lambda)$, ... $D_2(\lambda)$, $D_1(\lambda)$ son divisores de $D_{n_i-1}(\lambda)$ y, por ello, también son iguales a la unidad. Calculando los cocientes $D_{k+1}:D_k$ vemos que los factores invariantes de la matriz (12) serán $1, \ldots, 1, (\lambda-f(\rho_i))^{n_i}$, debido a lo cual la matriz (12) tendrá sólo un divisor elemental $(\lambda-f(\rho_i))^{n_i}$. De aquí se desprende el teorema siguiente:

TEOREMA 2. *Sea A una matriz de valores propios $\rho_1, \ldots, \rho_s$ y sea $f(\lambda)$ una función tal que $f'(\rho_i) \neq 0$ $(i=1, \ldots, s)$. Entonces, si la matriz $f(A)$ existe, sus divisores elementales se pueden obtener sustituyendo cada uno de los divisores elementales $(\lambda-\rho_i)^{n_i}$ de la matriz A por la expresión $(\lambda-f(\rho_i))^{n_i}$.*

Por ejemplo, si A es una matriz regular y $f(\lambda) = \lambda^{-1}$, se tiene $f(A) = A^{-1}$ y $f'(\rho_i) = -\rho_i^{-2} \neq 0$. Luego, si sustituimos todo divisor

elemental $(\lambda-\rho_i)^{n_i}$ de la matriz A por la expresión $(\lambda-\rho_i^{-1})^{n_i}$, obtendremos el sistema de divisores elementales de la matriz inversa.

El teorema 2 permite determinar los divisores elementales de la matriz $f(A)$ pertenecientes a aquellos valores propios $f(\rho_i)$ para los cuales $f'(\rho_i) \neq 0$. No es difícil obtener la regla correspondiente también para el caso en que $f'(\rho_i) = 0$. Supongamos que para un valor propio de una matriz A se tiene

$$f'(\rho_i) = f''(\rho_i) = \ldots = f^{(k-1)}(\rho_i) = 0 \text{ y } f^{(k)}(\rho_i) \neq 0. \tag{14}$$

Queremos determinar en qué divisores elementales de la matriz $f(A)$ se transforma cada uno de los divisores elementales $(\lambda-\rho_i)^{n_i}$ de la matriz A. Es obvio que este problema se reduce al siguiente: determinar los divisores elementales de la matriz (6) en la condición (14). Para $k \geqslant n_i$ la matriz (6) resulta diagonal y sus divisores elementales serán $\lambda - f(\rho_i), \ldots, \lambda - f(\rho_i)$. El caso $k = 1$ ha sido examinado anteriormente: en este caso la matriz (12) tiene un único divisor elemental $(\lambda - f(\rho_i))^{n_i}$. Por esto sólo nos interesarán los valores de k comprendidos entre 1 y n_i.

Consideremos un espacio lineal auxiliar $\mathfrak{L}$ de dimensión n_i y de base $a_1, a_2, \ldots, a_{n_i}$. Pongamos para abreviar $\frac{1}{l!} f^{(l)}(\rho_i) = \alpha_{l+1}$ y $n_i = n$ e indiquemos por $\mathfrak{C}$ la aplicación lineal del espacio $\mathfrak{L}$ que tiene la matriz

$$C = f(B_i) - f(\rho_i) E_i.$$

Tenemos, por consiguiente,

$$\left.\begin{array}{ll} a_1\mathfrak{C} & = \alpha_{k+1}a_{k+1} + \alpha_{k+2}a_{k+2} + \ldots + \alpha_n a_n, \\ a_2\mathfrak{C} & = \alpha_{k+1}a_{k+2} + \ldots + \alpha_{n-1}a_n, \\ \ldots & \ldots\ldots\ldots\ldots\ldots\ldots \\ a_{n-k}\mathfrak{C} & = \alpha_{k+1}a_n, \\ a_l\mathfrak{C} & = 0 \qquad (l > n-k). \end{array}\right\} \tag{15}$$

Puesto que nos interesan los divisores elemantales de la aplicación $\mathfrak{C}$, tomaremos en $\mathfrak{L}$ otra base en la que la matriz de la aplicación $\mathfrak{C}$ tenga una forma más sencilla. Sea

$$e_i = \beta_{i1}a_1 + \beta_{i,\,i+1}a_{i+1} + \ldots + \beta_{in}a_n \quad (i = 1, \ldots, n);$$

entonces

$$e_i\mathfrak{C} = \beta_{i1}\alpha_{k+1}a_{i+k} + (\beta_{i1}\alpha_{k+2} + \beta_{i,\,i+1}\alpha_{k+1})\,a_{i+k+1} + \ldots$$

Escojamos los números β_{ij} de modo que se cumplan las relaciones

$$e_i\mathfrak{C} = e_{i+k},\ e_i\mathfrak{C} = 0 \quad (i = 1, \ldots, n-k;\ j > n-k). \tag{16}$$

Esto ofrece el siguiente sistema de ecuaciones respecto de β_{ij}:

$$\alpha_{k+1}\beta_{ii} = \beta_{i+k,\,i+k},\ \alpha_{k+1}\beta_{i,\,i+1} + \alpha_{k+2}\beta_{ii} = \beta_{i+k,\,i+k+1}, \ldots$$

Teniendo en cuenta la condición $\alpha_{k+1} \neq 0$, de aquí se pueden determinar sucesivamente los valores $\beta_{ii}, \beta_{i,\,i+1}, \ldots$, expresándolos en términos de los valores de β con mayor primer índice. Para los valores β_{ij}, donde $i \geqslant n-k$, no se obtiene ninguna ecuación, de modo que estos valores se pueden escoger arbitrariamente. En particular, tomando $\beta_{n-k,\,n-k} = \ldots = \beta_{nn} = 1$, obtendremos para los demás coeficientes iniciales $\beta_{11}, \ldots, \beta_{n-k+1,\,n-k+1}$ valores nulos. Las ecuaciones (15) pueden ser resueltas entonces respecto de $a_1, a_2, \ldots, a_n$; por esto el sistema $e_1, e_2, \ldots, e_n$ será una base nueva de $\mathfrak{L}$ que satisface las

condiciones (16). Descompongamos los vectores $e_1, \ldots, e_n$ en los sistemas

$$e_1, e_{1+k}, e_{1+2k}, \ldots$$
$$e_2, e_{2+k}, e_{2+2k}, \ldots$$
$$\ldots\ldots\ldots\ldots$$
$$e_k, e_{2k}, e_{3k}, \ldots$$

y consideremos los subespacios $\mathfrak{L}_1, \ldots, \mathfrak{L}_k$, tendidos sobre estos sistemas. Las relaciones (16) muestran que $\mathfrak{L}_l$ es un subespacio invariante de base $e_l, e_{l+k}, e_{l+2k}, \ldots$ y que

$$\mathfrak{L} = \mathfrak{L}_1 + \mathfrak{L}_2 + \ldots + \mathfrak{L}_k. \tag{17}$$

En virtud de los resultados del p. 12.3, de aquí se desprende que la matriz de la aplicación $\mathcal{G}$ se descompone en k células de Jordan de divisores elementales $\lambda^{m_1}, \lambda^{m_2}, \ldots, \lambda^{m_k}$, donde $m_1, m_2, \ldots, m_k$ son las dimensiones de los subespacios $\mathfrak{L}_1, \mathfrak{L}_2, \ldots, \mathfrak{L}_k$. La matriz de la aplicación inicial $f(B_l)$ está ligada a la matriz C mediante la fórmula $f(B_l) = f(\rho_l) E_l + C$ y, por lo tanto, los divisores elementales de la matriz $f(B_l)$ son $(\lambda - f(\rho_l))^{m_1}, \ldots, (\lambda - f(\rho_l))^{m_k}$. Si empleamos el símbolo $[\alpha]$ para indicar el mayor número entero que no sobrepasa a α, obtendremos para $m_1, \ldots, m_k$ las expresiones siguientes:

$$m_1 = \left[\frac{n-1}{k}\right] + 1, \quad m_2 = \left[\frac{n-2}{k}\right] + 1, \quad \ldots, \quad m_k = \left[\frac{n-k}{k}\right] + 1 = \left[\frac{n}{k}\right].$$

Por consiguiente, *si la matriz A tiene un divisor elemental $(\lambda - \rho)^m$ y si $f'(\rho) = \ldots = f^{(k-1)}(\rho) = 0$ y $f^{(k)}(\rho) \neq 0$, entonces al pasar de A a $f(A)$ este divisor elemental se descompone en los divisores elementales $(\lambda - f(\rho))^{m_1}, \ldots, (\lambda - f(\rho))^{m_k}$, donde*

$$m_1 = \left[\frac{m-1}{k}\right] + 1, \quad m_2 = \left[\frac{m-2}{k}\right] + 1, \quad \ldots, \quad m_k = \left[\frac{m-k}{k}\right] + 1.$$

Consideremos un ejemplo. Supongamos que A tiene los divisores elementales $(\lambda - 1)^6$ y $(\lambda + 2)^3$; es preciso determinar los divisores elementales de la matriz $A^6 - 3A^4 + 3A^2 - E$. En estas condiciones tenemos $f(\lambda) = \lambda^6 - 3\lambda^4 + 3\lambda^2 - 1 = (\lambda - 1)^3 (\lambda + 1)^3$, $f'(-2) \neq 0$, $f'(1) = f''(1) = 0$ y $f'''(1) \neq 0$. Al pasar a $f(A)$ el divisor elemental $(\lambda + 2)^3$ se convierte en $(\lambda - f(-2))^3 = (\lambda - 27)^3$, mientras que el divisor elemental $(\lambda - 1)^6$ se descompone en λ^3, λ^2 y λ^1. Por consiguiente, los divisores elementales de la matriz $f(A)$ serán λ^3, λ^2, λ^1 y $(\lambda - 27)^3$.

16.5. Series de potencias. Una sucesión de matrices cuadradas

$$A_1, A_2, \ldots, A_m, A_{m+1}, \ldots \tag{18}$$

de un mismo orden se llama *convergente* hacia la matriz A, si los elementos de las matrices (18) que aparecen en la intersección de una columna y de una fila dadas convergen hacia el elemento correspondiente de la matriz A. De esta definición se desprende directamente que si las matrices A_m y B_m convergen con el crecimiento de m hacia A y B, respectivamente, las matrices $A_m + B_m$ y $A_m B_m$ convergen hacia $A + B$ y AB. En particular, si T es una matriz constante y la matriz A_m converge hacia A, resulta que $T^{-1} A_m T$ tendrá como límite a la matriz $T^{-1} A T$. Además, si

$$A_m = A_m^{(1)} + A_m^{(2)} + \ldots + A_m^{(s)} \quad (m = 1, 2, \ldots),$$

donde los órdenes de las células no dependen de m, la matriz A_m converge con el crecimiento de m hacia un límite determinado si, y sólo si, cada una de las células $A_m^{(l)}$ converge hacia un límite.

La última observación permite resolver fácilmente el problema de convergencia de las así llamadas series de potencias de matrices. Sea

$$\alpha_0+\alpha_1\lambda+\alpha_2\lambda^2+\ldots+\alpha_m\lambda^m+\ldots \tag{19}$$

una serie formal respecto de la variable λ. La expresión

$$\alpha_0 E+\alpha_1 A+\alpha_2 A^2+\ldots+\alpha_m A^m+\ldots \tag{20}$$

se llama *serie de potencias* correspondiente de la matriz A y el polinomio

$$f_n(A)=\alpha_0 E+\alpha_1 A+\ldots+\alpha_n A^n$$

se llama *n-ésima suma inicial* de esta serie. Se dice que la serie (20) *converge*, si la sucesión de las sumas iniciales $f_1(A), \ldots, f_m(A), \ldots$ tiene límite; en el caso de su existencia este límite se denomina *suma* de la serie (20).

Reduzcamos la matriz A a la forma normal

$$T^{-1}AT=B=B_1+B_2+\ldots+B_l,$$

donde $B_1, \ldots, B_l$ son células de Jordan. Hemos visto más arriba que la convergencia de la sucesión $f_m(A)$ equivale a la convergencia de la sucesión $T^{-1}f_m(A)T$ $(m=1, 2, \ldots)$. Pero

$$T^{-1}f_m(A)T=f_m(T^{-1}AT)=f_m(B)=f_m(B_1)+\ldots+f_m(B_l),$$

es decir, el problema acerca de la convergencia de la serie (20) equivale al siguiente: ¿bajo qué condiciones converge esta serie para las células de Jordan $B_1, \ldots, B_l$? Consideremos una de estas células, por ejemplo, la célula B_l. Sea $(\lambda-\rho_l)^{n_l}$ el divisor elemental que le corresponde. Según la fórmula (3) tenemos

$$f_m(B_l)=\begin{bmatrix} f_m(\rho_l) & \frac{1}{1!}f_m'(\rho_l) & \ldots & \frac{1}{(n_l-1)!}f_m^{(n_l-1)}(\rho_l) \\ & f_m(\rho_l) & \ldots & \frac{1}{(n_l-2)!}f_m^{(n_l-2)}(\rho_l), \\ & \ldots\ldots\ldots\ldots & & \ldots\ldots \\ & & & f_m(\rho_l) \end{bmatrix},$$

por consiguiente, $f_m(B_l)$ converge hacia un límite con el crecimiento de m si, y sólo si, $f_m(\rho_l), f_m'(\rho_l), \ldots, f_m^{(n_l-1)}(\rho_l)$ convergen hacia unos límites, es decir, si en el punto ρ_l convergen tanto la serie (19) como las series que se obtienen de ella derivándola término por término n_l-1 veces sucesivas. De la teoría de las funciones analíticas se sabe que todas estas series convergen indudablemente, si ρ_l pertenece al interior del círculo de convergencia de la serie (19) o si ρ_l pertenece a la circunferencia del círculo de convergencia y la (n_l-1)-ésima derivada de la serie (19) converge en el punto ρ_l. Es decir, hemos demostrado el teorema siguiente:

TEOREMA 3. *Para que una serie de potencias de una matriz A converja es necesario y suficiente que todo valor propio ρ_l de la matriz A se halle en el interior del círculo de convergencia de la correspondiente serie de potencias $f(\lambda)$ o que se halle sobre la circunferencia del círculo de convergencia, pero con la particularidad de que la serie, que se obtiene derivando n_l-1 veces la serie de $f(\lambda)$, converja en el punto ρ_l, donde n_l es el grado del mayor divisor elemental correspondiente a ρ_l.*

16.6. Matrices conmutables con una matriz dada. Dos matrices A y B se llaman *conmutables*, si $AB=BA$. Toda matriz es conmutable con sí misma y con la matriz unidad. Además, si A es

conmutable con las matrices B y C, las igualdades

$$A \cdot BC = BAC = BC \cdot A,$$

$$A(\alpha B + \beta C) = \alpha AB + \beta AC = \alpha BA + \beta CA = (\alpha B + \beta C) A$$

muestran que A es conmutable con el producto y las combinaciones lineales de las mismas. Luego, si B es conmutable con A, resulta que B es conmutable con cualquier polinomio en A. En particular, A es conmutable con los polinomios en A y cualesquiera dos polinomios en A son conmutables.

Según el p. 16.3, los valores de las funciones escalares de una matriz A pueden ser representados mediante polinomios en A. Los polinomios en A son conmutables con toda matriz B que sea conmutable con A. Por consiguiente, *el valor de toda función escalar de una matriz A es conmutable con todas las matrices que son conmutables con A.*

La relación $\alpha E \cdot P = P \cdot \alpha E$ muestra que *las matrices αE son conmutables con todas las matrices del mismo orden.* La recíproca también es válida:

Si una matriz cuadrada A de orden n es conmutable con todas las matrices de orden n, la matriz A es de la forma αE.

Indiquemos por α_{ij} los elementos de la matriz A. Sea P la matriz que contiene la unidad en la p-ésima fila y en la q-ésima columna, mientras que todas sus demás posiciones están ocupadas por ceros. Realizando la multiplicación directa obtenemos

$$AP = \begin{bmatrix} 0 & \dots & \alpha_{1p} & \dots & 0 \\ 0 & \dots & \alpha_{2p} & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \dots & \alpha_{np} & \dots & 0 \end{bmatrix} \quad \text{y} \quad PA = \begin{bmatrix} 0 & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \alpha_{q1} & \alpha_{q2} & \dots & \alpha_{qn} \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & 0 \end{bmatrix},$$

donde aparecen escritas la q-ésima columna de la primera matriz y la p-ésima fila de la segunda. Como, por hipótesis, $AP = PA$, resulte que $\alpha_{pp} = \alpha_{qq}$ y que $\alpha_{pq} = 0$, si $p \neq q$. Puesto que p y q son arbitrarios, esto significa que A es diagonal y que todos sus elementos diagonales son iguales.

Consideremos ahora un problema más complejo: *hallar todas las matrices conmutables con una matriz dada A.*

Para resolverlo, reduciremos A a la forma normal de Jordan

$$T^{-1}AT = B = B_1 \dotplus B_2 \dotplus \dots \dotplus B_s, \qquad (21)$$

donde $B_1, \dots, B_s$ son células de Jordan. Si la matriz X es conmutable con B, es obvio que $Y = TXT^{-1}$ es conmutable con A y, recíprocamente, si Y conmuta con A, también $T^{-1}YT = X$ conmuta con B. Por lo tanto, todo el problema se reduce a determinar las matrices X que conmutan con la matriz B que tiene la forma

normal (21). De acuerdo con (21) descompongamos en células también la matriz X:

$$X = \begin{vmatrix} X_{11} & X_{12} & \dots & X_{1s} \\ X_{21} & X_{22} & \dots & X_{2s} \\ \dots & \dots & \dots & \dots \\ X_{s1} & X_{s2} & \dots & X_{ss} \end{vmatrix}.$$

La condición $BX = XB$ lleva a las igualdades

$$B_p X_{pq} = X_{pq} B_q \quad (p, q = 1, \dots, s). \tag{22}$$

Vemos que para cada una de las células X_{pq} se obtiene una sola igualdad (22) de la cual debemos determinar los elementos de X_{pq}.

Supongamos que los órdenes de las matrices B_p y B_q son k y m, respectivamente, y que sus valores propios son ρ y σ. En este caso X_{pq} será una matriz rectangular de k filas y de m columnas. Indiquemos por ξ_{ij} $(i = 1, \dots, k;\ j = 1, \dots, m)$ los elementos de la matriz X_{pq} y representemos la relación (22) en forma detallada

$$\begin{bmatrix} \rho & 1 & 0 & \dots & 0 \\ & \rho & 1 & \dots & 0 \\ & & & \dots & \\ & & & & \rho \end{bmatrix} \begin{bmatrix} \xi_{11} & \xi_{12} & \dots & \xi_{1m} \\ \xi_{21} & \xi_{22} & \dots & \xi_{2m} \\ \dots & \dots & \dots & \dots \\ \xi_{k1} & \xi_{k2} & \dots & \xi_{km} \end{bmatrix} - \begin{bmatrix} \xi_{11} & \xi_{12} & \dots & \xi_{1m} \\ \xi_{21} & \xi_{22} & \dots & \xi_{2m} \\ \dots & \dots & \dots & \dots \\ \xi_{k1} & \xi_{k2} & \dots & \xi_{km} \end{bmatrix} \begin{bmatrix} \sigma & 1 & 0 & \dots & 0 \\ & \sigma & 1 & \dots & 0 \\ & & & \dots & \\ & & & & \sigma \end{bmatrix}.$$

Realizando aquí la multiplicación y comparando el elemento que se obtiene en la i-ésima fila y j-ésima columna del primer miembro con el elemento correspondiente del segundo miembro, llegamos a las ecuaciones

$$\rho\xi_{ij} + \xi_{i+1,\,j} = \xi_{i,\,j-1} + \sigma\xi_{ij} \quad (i \neq k,\ j \neq 1), \tag{23}$$

$$\rho\xi_{kj} = \xi_{k,\,j-1} + \sigma\xi_{kj} \quad (i = k,\ j \neq 1), \tag{24}$$

$$\rho\xi_{k1} = \sigma\xi_{k1} \quad (i = k,\ j = 1). \tag{25}$$

Si $\rho \neq \sigma$, de (25) se deduce que $\xi_{k1} = 0$; entonces de (24) obtenemos sucesivamente $\xi_{k2} = \dots = \xi_{km} = 0$ y de (23) concluimos que todos los demás ξ_{ij} son también iguales a cero. Por consiguiente, para $\rho \neq \sigma$ tenemos $X_{pq} = O$.

Consideremos el caso $\rho = \sigma$. Las ecuaciones (23), (24) y (25) se convierten entonces en

$$\xi_{i+1,\,j} = \xi_{i,\,j-1} \quad (i = 1, \dots, k-1;\ j = 2, \dots, m), \tag{26}$$

$$\xi_{k,\,j-1} = 0 \quad (j = 2, \dots, m). \tag{27}$$

Sea $k \geqslant m$; poniendo $\xi_{11} = \xi_1$, $\xi_{12} = \xi_2$, $\dots$, $\xi_{1m} = \xi_m$, reduciremos las ecuaciones (26) y (27) a las ecuaciones

$$\xi_{ij} = \xi_{j-i+1} \quad (i \leqslant j),$$

$$\xi_{ij} = 0 \quad (i > j),$$

de donde resulta que la matriz X_{pq} tiene la forma *triangular lineal*

$$X_{pq}=\begin{bmatrix} \xi_1 & \xi_2 & \xi_3 & \cdots & \xi_m \\ & \xi_1 & \xi_2 & \cdots & \xi_{m-1} \\ & & \cdots & \cdots & \cdots \\ & & & & \xi_1 \end{bmatrix} \tag{28}$$

para $k=m$ y, respectivamente, la forma

$$X_{pq}=\begin{bmatrix} \xi_1 & \xi_2 & \xi_3 & \cdots & \xi_m \\ & \xi_1 & \xi_2 & \cdots & \xi_{m-1} \\ & & \cdots & \cdots & \cdots \\ & & & & \xi_1 \\ 0 & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

para $k>m$. Siendo $k<m$ y tomando $\xi_{1,\,m-k+1}=\xi_1$, $\xi_{1,\,m-k+2}=\xi_2$, ..., $\xi_{1m}=\xi_k$, reduciremos las ecuaciones (23), (24) y (25) a las ecuaciones

$$\xi_{ij}=\xi_{j-i-m+k+1} \quad (j-i\geqslant m-k),$$
$$\xi_{ij}=0 \quad (j-i<m-k),$$

que significan que la matriz X_{pq} es de la forma

$$X_{pq}=\begin{bmatrix} 0 & \cdots & 0 & \xi_1 & \xi_2 & \cdots & \xi_k \\ 0 & \cdots & 0 & 0 & \xi_1 & \cdots & \xi_{k-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & \xi_1 \end{bmatrix}.$$

Recíprocamente, si las células de la matriz X tienen la forma señalada, las ecuaciones (23), (24) y (25) se satisfacen y, por consiguiente, X conmuta con B.

Por ejemplo, si

$$B=\begin{bmatrix} \rho & 1 & 0 \\ & \rho & 1 \\ & & \rho \end{bmatrix}+\begin{bmatrix} \rho & 1 \\ & \rho \end{bmatrix}+\begin{bmatrix} \sigma & 1 & 0 \\ & \sigma & 1 \\ & & \sigma \end{bmatrix} \quad (\rho\neq\sigma), \tag{29}$$

las matrices que conmutan con B son de la forma

$$X=\left[\begin{array}{ccc:cc} \alpha_0 & \alpha_1 & \alpha_2 & \gamma_2 & \gamma_1 \\ & \alpha_0 & \alpha_1 & 0 & \gamma_0 \\ & & \alpha_0 & 0 & 0 \\ \hdashline 0 & \delta_0 & \delta_1 & \beta_0 & \beta_1 \\ 0 & 0 & \delta_0 & 0 & \beta_0 \end{array}\right]+\begin{bmatrix} \lambda_0 & \lambda_1 & \lambda_2 \\ & \lambda_0 & \lambda_1 \\ & & \lambda_0 \end{bmatrix},$$

donde α_i, β_i, γ_i y δ_i son números arbitrarios.

Este resultado se hace sumamente sencillo para matrices de tipo $\rho_1E_1+\rho_2E_2+\ldots+\rho_sE_s$, donde $E_1, \ldots, E_s$ son matrices unidades

y todos los números $\rho_1, \ldots, \rho_s$ son diferentes. En este caso las células X_{pq} son nulas para $p \neq q$ y X adquiere la forma celular diagonal.

En particular, si B es diagonal y sus elementos diagonales son distintos, las matrices conmutables con B son también diagonales.

16.7. Matrices que conmutan con matrices conmutables. Los polinomios en una matriz A poseen la propiedad específica de que, además de conmutar con la propia matriz A, conmutan con cualquier matriz X que es conmutable con A. Resulta que esta propiedad es característica para los polinomios en A.

TEOREMA 4. *Si la matriz C conmuta con todas las matrices que conmutan con B, C es un polinomio en B.*

Es obvio que basta realizar la demostración para el caso en que B tiene la forma normal de Jordan; sea, pues,

$$B = B_1 \dotplus B_2 \dotplus \ldots \dotplus B_s,$$

donde $B_1, \ldots, B_s$ son células de Jordan. Las matrices auxiliares

$$X = \alpha_1 E_1 + \alpha_2 E_2 \dotplus \ldots \dotplus \alpha_s E_s,$$

donde $\alpha_1, \ldots, \alpha_s$ son números arbitrarios y $E_1, \ldots, E_s$ son matrices unidades, conmutan indudablemente con B y, por ello, las matrices X conmutan también con C. De aquí se deduce, de acuerdo con lo expuesto anteriormente, que C se descompone en células:

$$C = C_1 \dotplus C_2 \dotplus \ldots \dotplus C_s; \tag{30}$$

además de la condición $CB = BC$ resulta que estas células tienen la forma lineal triangular (28). Sea ahora X una matriz arbitraria que conmuta con B. La forma general de la matriz X ha sido determinada en el punto anterior. Según las condiciones del teorema que estamos demostrando, la matriz C debe conmutar con X. Representando X en la forma celular veremos que la igualdad $CX = XC$ equivale a las relaciones

$$C_p X_{pq} = X_{pq} C_q \qquad (p, q = 1, \ldots, s). \tag{31}$$

Si las células correspondientes B_p y B_q tienen diferentes valores propios, nada resulta de (31) ya que en este caso $X_{pq} = O$. Por esto aceptaremos que los valores propios de las células B_p y B_q coinciden. Sea

$$C_p = \begin{bmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_k \\ & \alpha_1 & \ldots & \alpha_{k-1} \\ & & \ldots\ldots & \\ & & & \alpha_1 \end{bmatrix} \quad \text{y} \quad C_q = \begin{bmatrix} \beta_1 & \beta_2 & \ldots & \beta_m \\ & \beta_1 & \ldots & \beta_{m-1} \\ & & \ldots\ldots & \\ & & & \beta_1 \end{bmatrix}.$$

Supongamos, para concretar, que $k < m$. Entonces la relación (31)

se convierte en

$$\begin{bmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_k \\ & \alpha_1 & \dots & \alpha_{k-1} \\ & & \dots\dots & \\ & & & \alpha_1 \end{bmatrix} \begin{bmatrix} 0 & \dots & 0 & \xi_1 & \xi_2 & \dots & \xi_k \\ 0 & \dots & 0 & 0 & \xi_1 & \dots & \xi_{k-1} \\ & & & \dots\dots\dots\dots & & & \\ 0 & \dots & 0 & 0 & 0 & \dots & \xi_1 \end{bmatrix} =$$

$$= \begin{bmatrix} 0 & \dots & 0 & \xi_1 & \xi_2 & \dots & \xi_k \\ 0 & \dots & 0 & 0 & \xi_1 & \dots & \xi_{k-1} \\ & & & \dots\dots\dots\dots & & & \\ 0 & \dots & 0 & 0 & 0 & \dots & \xi_1 \end{bmatrix} \begin{bmatrix} \beta_1 & \beta_2 & \dots & \beta_m \\ & \beta_1 & \dots & \beta_{m-1} \\ & & & \dots\dots \\ & & & \beta_1 \end{bmatrix},$$

donde $\xi_1, \xi_2, \dots, \xi_m$ son números arbitrarios. Multiplicando en el primer y en el segundo miembros la primera fila por la última columna e igualando los resultados, obtenemos

$$\alpha_1\xi_k + \alpha_2\xi_{k-1} + \dots + \alpha_k\xi_1 = \beta_1\xi_k + \beta_2\xi_{k-1} + \dots + \beta_k\xi_1,$$

de donde, debido a la arbitrariedad de los números $\xi_1, \dots, \xi_k$ resulta que

$$\alpha_1 = \beta_1, \quad \alpha_2 = \beta_2, \quad \dots, \quad \alpha_k = \beta_k. \tag{32}$$

Es fácil comprobar que semejantes igualdades se obtienen también en el caso en que $k \geqslant m$. Las igualdades (30), (31) y (32) constituyen un sistema completo de condiciones a las que debe someterse la matriz incógnita C. Para hacer estas condiciones más claras, procederemos del modo siguiente. Coloquemos las células de Jordan de la matriz B de manera que las células con los valores propios iguales estén al lado una de otra. Sea, por ejemplo,

$$B = (B_1 \dotplus \dots \dotplus B_{m_1}) \dotplus (B_{m_1+1} \dotplus \dots \dotplus B_{m_2}) \dotplus \dots$$
$$\dots \dotplus (B_{m_{t-1}+1} \dotplus \dots \dotplus B_s),$$

donde en cada uno de los paréntesis figuran células con valores propios iguales. Indicando las sumas de estos paréntesis por $B^{(1)}, \dots B^{(t)}$, la matriz B quedará dividida en células mayores que denominaremos bloques. De acuerdo con esto también las matrices X y C quedarán divididas en bloques respectivos. Los resultados del punto anterior muestran que todos los bloques no diagonales de la matriz X son iguales a cero; en cuanto a los bloques diagonales de X, éstos tienen la estructura descrita en aquel punto. Las condiciones obtenidas en el punto presente para la matriz C muestran que sus bloques diagonales también se descomponen, igual que en el caso de B, en células, con la particularidad de que las células de la matriz C tienen una forma triangular especial. Las igualdades (32) significan que en las células de la matriz C, pertenecientes a un mismo bloque, los elementos que figuran en una misma línea paralela a la diagonal principal son iguales.

Sea, por ejemplo, B la matriz (29) considerada a título de ejemplo al final del punto anterior. Nuestros resultados muestran que la matriz C que conmuta con toda matriz conmutable con B será de la forma

$$C = \left[\begin{array}{ccccc|ccc} \alpha_0 & \alpha_1 & \alpha_2 & & & & & \\ & \alpha_0 & \alpha_1 & & & & & \\ & & \alpha_0 & & & & & \\ & & & \alpha_0 & \alpha_1 & & & \\ & & & & \alpha_0 & & & \\ \hline & & & & & \lambda_0 & \lambda_1 & \lambda_2 \\ & & & & & & \lambda_0 & \lambda_1 \\ & & & & & & & \lambda_0 \end{array}\right]. \tag{33}$$

Resta probar que C se puede representar en forma de un polinomio en B. Haremos esto para el caso particular en que B es de la forma (29) y, por consiguiente, C tiene la forma (33); en el caso general los razonamientos son los mismos. Es decir, debemos probar que la matriz (33) es un polinomio en la matriz (29). Busquemos un polinomio $f(\lambda)$ que satisfaga las condiciones

$$f(\rho) = \alpha_0, \quad f'(\rho) = \alpha_1, \quad f''(\rho) = \alpha_2,$$
$$f(\sigma) = \lambda_0, \quad f'(\sigma) = \lambda_1, \quad f''(\sigma) = \lambda_2.$$

Según el p. 16.3 existe un polinomio de esta índole. Aplicando la regla (3) del p. 16.1 veremos que $f(B) = C$.

Ejemplos y problemas

1. Hállese A^n, si

$$A = \begin{bmatrix} 1 & 4 & 2 \\ 0 & -3 & -2 \\ 0 & 4 & 3 \end{bmatrix}.$$

2. Hállense *todas* las soluciones de la ecuación $X^2 = A$, donde A tiene el mismo valor que en el problema anterior. ¿Cuáles de estas soluciones serán polinomios en A? Calcúlense $\operatorname{sen} \pi A$, e^A y $\cos \pi A$.

3. Los divisores elementales de una aplicación $\mathcal{A}$ son iguales a λ^5, $(\lambda - \pi)^3$ y $\lambda + \frac{\pi}{2}$. Calcúlense los divisores elementales de las aplicaciones $\cos \mathcal{A}$ y $\operatorname{sen} \mathcal{A}$.

4. Si A y B son matrices que conmutan, se tiene $e^A e^B = e^{A+B}$.

5. ¿Para qué matrices A tiene solución la ecuación $A = e^X$?

6. Si $f(\lambda)$ indica la suma de la serie de potencias (19), la suma de la serie de potencias (20), si es que esta serie converge, será igual a $f(A)$.

7. Hállese la forma general de la matriz que conmuta con una matriz normal de Jordan A, si los divisores elementales de A son $(\lambda - 1)^3$, $(\lambda - 1)^2$, $\lambda - 1$, $(\lambda + 2)^2$ y $\lambda + 2$.

8. Supongamos que todos los factores invariantes de la matriz $\lambda E - A$, salvo el último, son iguales a 1. Entonces toda matriz conmutable con A es un polinomio en A.

Capítulo V

Espacios unitarios y euclídeos

Los espacios lineales que hemos estudiado en los capítulos anteriores han resultado ser, en determinado sentido, más pobres en conceptos y propiedades que nuestro espacio corriente. En la teoría general de los espacios lineales no han quedado reflejados conceptos como la longitud de un segmento, la magnitud del ángulo y el producto escalar que desempeñan un papel primordial en la geometría. Por esto, si queremos que la teoría general abarque todas las propiedades más esenciales del espacio corriente, debemos introducir, además de las operaciones de adición de vectores y de multiplicación de los mismos por números, la operación de multiplicación escalar. En este capítulo se estudian precisamente las propiedades de los vectores pertenecientes a espacios provistos del producto escalar.

En este capítulo el cuerpo principal es de carácter muy especial: es el cuerpo de los números reales en el caso de espacios euclídeos y es el cuerpo de los números complejos en el caso de espacios unitarios.

§ 17. Espacios unitarios

17.1. Axiomática y ejemplos. Sea $\mathfrak{R}$ nuestro espacio corriente cuyos vectores son los segmentos orientados que parten de un punto inicial O. Se llama producto escalar (a, b) de los vectores a y b el producto de las longitudes de a y de b por el coseno del ángulo que forman estos vectores. De aquí se desprenden directamente las conocidas propiedades del producto escalar:

(a) $(a, b) = (b, a)$;
(b) $(\alpha a, b) = \alpha (a, b)$;
(c) $(a + b, c) = (a, c) + (b, c)$;
(d) si $a \neq o$, se tiene $(a, a) > 0$.

Tomemos en el espacio $\mathfrak{R}$ un sistema de coordenadas formado por tres cualesquiera vectores e_1, e_2 y e_3, perpendiculares dos a dos, de longitud 1. Entonces todo vector a admite una representación única de la forma

$$a = \alpha_1 e_1 + \alpha_2 e_2 + \alpha_3 e_3,$$

donde α_1, α_2 y α_3 son las longitudes de las proyecciones del vector a sobre los ejes coordenados, tomadas con signo adecuado. Si

$$b = \beta_1 e_1 + \beta_2 e_2 + \beta_3 e_3$$

es otro vector cualquiera, resulta de la definición del producto escalar y de las propiedades (b) y (c) que

$$(a, b) = \alpha_1\beta_1 + \alpha_2\beta_2 + \alpha_3\beta_3. \qquad (1)$$

El espacio $\mathfrak{R}$ es real. Esto se expresa en que las proyecciones, las longitudes y los productos escalares de los vectores son números reales. Sin embargo, en algunos casos surge la necesidad de considerar vectores de proyecciones complejas. A primera vista parece natural tomar de nuevo la expresión (1) para el producto escalar de vectores con coordenadas complejas α_1, α_2, α_3 y β_1, β_2, β_3. En algunos casos se procede precisamente de este modo. El espacio que así resulta se denomina *espacio euclídeo complejo*. Por desgracia, el producto escalar pierde entonces muchas propiedades importantes y entre ellas la propiedad (d) de importancia primordial. En efecto, para el vector

$$a = 3e_1 + 4e_2 + 5ie_3 \quad (i = \sqrt{-1})$$

de la fórmula (1) resulta

$$(a, a) = 9 + 16 + 25i^2 = 0$$

contrariamente a la propiedad (d). Para evitar este inconveniente, en lugar de la expresión (1) se toma como definición del producto escalar de vectores complejos la expresión

$$(a, b) = \alpha_1\bar{\beta}_1 + \alpha_2\bar{\beta}_2 + \alpha_3\bar{\beta}_3, \qquad (2)$$

donde la raya superior significa que ha de pasarse a los números complejos conjugados. En el caso en que los vectores a y b son reales, tenemos $\beta_j = \bar{\beta}_j$ y la expresión (2) coincide con (1). Por consiguiente, la nueva definición (2) es una generalización de la anterior. Por otra parte, con la nueva definición la propiedad (d) se cumple sin duda alguna, ya que de (2) resulta:

$$(a, a) = \alpha_1\bar{\alpha}_1 + \alpha_2\bar{\alpha}_2 + \alpha_3\bar{\alpha}_3 = |\alpha_1|^2 + |\alpha_2|^2 + |\alpha_3|^2,$$

donde $|\alpha_j|$ es el módulo del número α_j. Es fácil ver que las propiedades (b) y (c) también se verifican. En cuanto a la propiedad

(a), toma una forma distinta en el caso de vectores complejos. Efectivamente, de (2) se tiene

$$(b, a) = \beta_1\bar{\alpha}_1 + \beta_2\bar{\alpha}_2 + \beta_3\bar{\alpha}_3 = \overline{\bar{\beta}_1\alpha_1 + \bar{\beta}_2\alpha_2 + \bar{\beta}_3\alpha_3},$$

es decir,

$$(a^*) \qquad (b, a) = \overline{(a, b)}.$$

El espacio de vectores complejos en el que el producto escalar se calcula mediante la fórmula (2) se denomina unitario. La fórmula (a*) muestra que las propiedades del espacio unitario difieran, en general, de las propiedades del espacio corriente. No obstante, estas diferencias son de poca importancia. En todo caso, el espacio unitario se aproxima más por sus propiedades al espacio corriente que el espacio euclídeo complejo mencionado anteriormente.

Los razonamientos que hemos expuesto no pueden calificarse de totalmente precisos. Además, hemos considerado el caso de un espacio de tres dimensiones. Por esto debemos dar ahora una definición totalmente rigurosa de los espacios unitarios que sea válida también para espacios de cualquier dimensión.

En la teoría general de espacios lineales hemos realizado casi toda la exposición aceptando que el cuerpo principal K es totalmente arbitrario. En el capítulo presente K será o bien el cuerpo de todos los números complejos o bien el cuerpo de todos los números reales.

Un espacio lineal $\mathfrak{L}$ sobre un campo K se llama *unitario*, si a todo par de vectores a y b de $\mathfrak{L}$ tomados en un orden determinado corresponde un número de K llamado *producto escalar* (a, b) del vector a por el vector b que posee las propiedades siguientes:

1° $(a, b) = \overline{(b, a)}$;
2° $(\alpha a, b) = \alpha (a, b)$;
3° $(a + b, c) = (a, c) + (b, c)$;
4° *si* $a \neq o$, *se tiene* $(a, a) > 0$.

En el caso en que el campo principal K es el cuerpo de los números reales, el espacio unitario $\mathfrak{L}$ se denomina *espacio unitario real* o simplemente *espacio euclídeo real*. En este caso la expresión $\overline{(a, b)}$ coincide, obviamente, con la expresión (a, b) y el axioma 1° adquiere una forma más sencilla: $(a, b) = (b, a)$.

Si el campo K es el cuerpo de los números complejos, el espacio $\mathfrak{L}$ se llama *espacio complejo unitario*. En lo sucesivo las propiedades de los espacios euclídeos reales y las propiedades de los espacios complejos unitarios serán examinadas, en la mayoría de los casos, conjuntamente y por espacio unitario se comprenderá, de acuerdo con la definición, o bien el espacio unitario real o bien el espacio unitario complejo.

Notemos también que en la definición de los espacios unitarios no se exige que el espacio sea de dimensión finita. Por esto cabe hablar también de espacios unitarios de dimensión infinita. Aun cuando algunas propiedades de los espacios unitarios no dependen de la dimensión de los mismos, nos limitaremos a considerar, mientras que no se diga lo contrario, solamente espacios de dimensión finita. La teoría de espacios de dimensión infinita entra de lleno en una disciplina matemática especial que es el Análisis funcional.

De las propiedades 1° y 2° resulta

$$(\alpha a, \beta b) = \alpha (a, \beta b) = \alpha \overline{(\beta b, a)} = \alpha \bar{\beta} \overline{(b, a)} = \alpha \bar{\beta} (a, b),$$

es decir

$$(\alpha a, \beta b) = \alpha \bar{\beta} (a, b). \tag{3}$$

Análogamente de 1°, 2° y 3° se desprende que

$$(a, b + c) = \overline{(b + c, a)} = \overline{(b, a)} + \overline{(c, a)} = (a, b) + (a, c).$$

De aquí obtenemos, mediante el procedimiento corriente, la fórmula general

$$\left(\sum \alpha_j a_j, \sum \beta_k b_k\right) = \sum\sum \alpha_j \bar{\beta}_k (a_j, b_k). \tag{4}$$

Para $\alpha = 1$ y $\beta = 0$ de la relación (3) resulta

$$(a, 0) = (0, a) = 0.$$

Señalemos dos ejemplos. Consideremos el espacio lineal de filas de longitud n con elementos del campo K. Convengamos en llamar *producto escalar* de la fila $a = [\alpha_1, \ldots, \alpha_n]$ por la fila $b = [\beta_1, \ldots, \beta_n]$ la expresión

$$(a, b) = \alpha_1 \bar{\beta}_1 + \alpha_2 \bar{\beta}_2 + \ldots + \alpha_n \bar{\beta}_n. \tag{5}$$

De esta expresión se ve que

$$(a, a) = \alpha_1 \bar{\alpha}_1 + \ldots + \alpha_n \bar{\alpha}_n = |\alpha_1|^2 + \ldots + |\alpha_n|^2. \tag{6}$$

Puesto que los módulos $|\alpha_j|$ son números reales no negativos, la suma de sus cuadrados será un número real no negativo que será igual a cero sólo en el caso en que sean iguales a cero todos sus sumandos. Por consiguiente, la propiedad 4° aquí se cumple. Es obvio que las propiedades 1°, 2° y 3° también se cumplen, de modo que el espacio de filas con el producto escalar (5) es un espacio unitario de dimensión n sobre el campo K. Este ejemplo es de una importancia primordial ya que más adelante quedará demostrado que todos los espacios unitarios de dimensión n sobre el campo K son isomorfos.

Como ejemplo de espacio unitario de dimensión infinita puede servir el espacio Ω de todas las funciones continuas $f(t)$ de valores

complejos definidas sobre el segmento [0, 1]. La adición y la multiplicación por número de estas funciones se definen de modo corriente y el producto escalar de una función $f(t)$ por otra función $g(t)$ se define mediante la fórmula

$$(f, g) = \int_0^1 f(t)\,\overline{g(t)}\,dt. \tag{7}$$

Las propiedades de 1° a 4° se demuestran fácilmente y, por consiguiente, el espacio $\mathfrak{L}$ es unitario. En este ejemplo el campo principal es el cuerpo de todos los números complejos. Si nos limitamos a considerar solamente las funciones continuas de valores reales, se puede tomar como campo principal el cuerpo de los números reales. La fórmula (7) quedará sustituida entonces por la fórmula

$$(f, g) = \int_0^1 f(t)\,g(t)\,dt.$$

17.2. Longitud de un vector. El cuadrado escalar (a, a) de cualquier vector es, según el axioma 4°, un número real no negativo. El valor no negativo de la raíz cuadrada de este número se denomina *longitud* o *norma* del vector a y se designa por $\|a\|$. Es decir, por definición

$$\|a\| = \sqrt{(a, a)}.$$

De esta definición se ve directamente que *el vector nulo es el único vector cuya longitud es igual a cero*. Además, si α es un número, se tiene

$$\|\alpha a\| = \sqrt{(\alpha a, \alpha a)} = \sqrt{\alpha\bar{\alpha}\,(a, a)} = |\alpha|\,\sqrt{(a, a)}, \tag{8}$$

es decir, *al multiplicar un vector por un número su longitud se multiplica por el módulo de este número*. Un vector cuya longitud es igual a la unidad se llama *vector unidad* o *vector normalizado*. La igualdad (8) muestra que al multiplicar un vector no nulo por el número inverso de su longitud se obtiene un vector unidad. Esta operación se llama a veces *normalización* de un vector.

Pasamos a la demostración de una desigualdad importante que relaciona las longitudes de dos vectores con el valor del producto escalar de los mismos.

DESIGUALDAD DE CAUCHY–BUNIAKOVSKI. *Para cualesquiera dos vectores a y b de un espacio unitario es válida la desigualdad*

$$|(a, b)| \leqslant \|a\| \cdot \|b\|,$$

teniendo lugar la igualdad cuando, y sólo cuando, los vectores a y b son linealmente dependientes.

Para cualquier número λ se tiene en virtud del axioma 4°

$$(a-\lambda b,\ a-\lambda b)\geqslant 0, \tag{9}$$

de donde realizando la multiplicación obtenemos

$$(a,\ a)-\bar{\lambda}(a,\ b)-\lambda\overline{(a,\ b)}+\lambda\bar{\lambda}(b,\ b)\geqslant 0. \tag{10}$$

Si $b=o$, la desigualdad requerida se cumple de un modo trivial ya que ambos miembros suyos resultan ser iguales a cero. Supongamos, por ello, que $b\neq o$. Tomando en la desigualdad (10) en lugar de λ el número $\frac{(a,\ b)}{(b,\ b)}$ y multiplicando todos los miembros de la desigualdad por el número positivo $(b,\ b)$, obtenemos

$$(a,\ a)(b,\ b)-\overline{(a,\ b)}(a,\ b)-(a,\ b)\overline{(a,\ b)}+(a,\ b)\overline{(a,\ b)}\geqslant 0,$$

es decir,

$$(a,\ b)\overline{(a,\ b)}\leqslant(a,\ a)(b,\ b)$$

o

$$|(a,\ b)|\leqslant\|a\|\cdot\|b\|. \tag{11}$$

Luego, hemos demostrado la desigualdad de Cauchy—Buniakovski.

Si a y b son linealmente independientes, se tiene $a-\lambda b\neq o$ y en lugar de (9) podemos tomar la desigualdad estricta

$$(a-\lambda b,\ a-\lambda b)>0.$$

Pero entonces, podemos omitir en todas las desigualdades sucesivas el signo de igualdad y en lugar de (11) obtendremos la desigualdad

$$|(a,\ b)|<\|a\|\cdot\|b\|. \tag{12}$$

En cambio, si a y b son linealmente dependientes, por ejemplo, $a=\alpha b$, tenemos

$$|(a,\ b)|=|(\alpha b,\ b)|=|\alpha|(b,\ b)=\|a\|\cdot\|b\|.$$

Luego, hemos demostrado también la observación complementaria a la desigualdad de Buniakovski.

Veamos qué es lo que significa la desigualdad de Buniakovski en aquellos espacios concretos que hemos considerado anteriormente. Sea $\mathfrak{L}$ el espacio unitario de filas. Tomemos en el mismo unos vectores $a=[\alpha_1,\ \ldots,\ \alpha_n]$ y $b=[\beta_1,\ \ldots,\ \beta_n]$. Tenemos según la fórmula (6)

$$|(a,\ b)|=|\alpha_1\bar{\beta}_1+\alpha_2\bar{\beta}_2+\ldots+\alpha_n\bar{\beta}_n|,$$

$$\|a\|=\sqrt{|\alpha_1|^2+\ldots+|\alpha_n|^2}\quad\text{y}\quad\|b\|=\sqrt{|\beta_1|^2+\ldots+|\beta_n|^2}.$$

La desigualdad de Cauchy—Buniakovski significa, por consiguiente, que

$$\alpha_1\bar{\beta}_1+\ldots+\alpha_n\bar{\beta}_n\leqslant\sqrt{|\alpha_1|^2+\ldots+|\alpha_n|^2}\sqrt{|\beta_1|^2+\ldots+|\beta_n|^2},$$

donde α_j y β_j son números complejos cualesquiera.

Análogamente, si $\mathfrak{L}$ es el espacio de funciones mencionado anteriormente, la desigualdad de Buniakovski se convierte en la desigualdad

$$\left|\int_0^1 f(t)\,\overline{g(t)}\,dt\right|^2 \leqslant \int_0^1 |f(t)|^2\,dt \int_0^1 |g(t)|^2\,dt.$$

Volvamos a los espacios unitarios arbitrarios. La desigualdad de Cauchy—Buniakovski permite ahora demostrar fácilmente la siguiente proposición: *la longitud de una suma de vectores no sobrepasa la suma de las longitudes de los sumandos*. Es obvio que basta considerar el caso de dos sumandos solamente, ya que el caso general se deduce de éste por inducción. Tenemos

$$\|a+b\|^2 = (a+b,\ a+b) = (a,\ a)+(a,\ b)+$$
$$+\overline{(a,\ b)}+(b,\ b) = (a,\ a)+2\,\mathrm{Re}\,(a,\ b)+(b,\ b),$$

donde $\mathrm{Re}\,(a,\ b)$ es la parte real de $(a,\ b)$. Puesto que

$$\mathrm{Re}\,(a,\ b) \leqslant |(a,\ b)| \leqslant \|a\|\cdot\|b\|,$$

se tiene

$$\|a+b\|^2 \leqslant \|a\|^2+2\|a\|\cdot\|b\|+\|b\|^2 = (\|a\|+\|b\|)^2,$$

de donde resulta

$$\|a+b\| \leqslant \|a\|+\|b\|$$

que es lo que se quería demostrar.

La expresión $\|a-b\|$ suele llamarse a veces *distancia entre los vectores* a y b. Indicándola por $\rho(a,\ b)$ obtenemos las relaciones siguientes

1) $\rho(a,\ a)=0$; $\rho(a,\ b)>0$, si $a\neq b$;
2) $\rho(a,\ b)=\rho(b,\ a)$;
3) $\rho(a,\ b)+\rho(b,\ c)\geqslant\rho(a,\ c)$.

La demostración de las mismas es evidente; por ejemplo, la última resulta de

$$\rho(a,\ c) = \|a-c\| = \|(a-b)+(b-c)\| \leqslant$$
$$\leqslant \|a-b\|+\|b-c\| = \rho(a,\ b)+\rho(b,\ c).$$

17.3. Sistemas ortonormales. Unos vectores a y b de un espacio unitario $\mathfrak{L}$ se llaman *ortogonales* si el producto escalar de a por b es igual a cero. Si $\mathfrak{L}$ es el espacio corriente, el concepto de ortogonalidad coincide con el concepto de perpendicularidad. Por esto la ortogonalidad puede ser considerada como una generalización del concepto de perpendicularidad.

Del axioma 1° se deduce que la relación de ortogonalidad es simétrica: si el vector a es ortogonal a b, el vector b es ortogonal

a a. Es obvio que *el vector nulo es ortogonal a cualquier vector del espacio y es el único vector que posee esta propiedad.*

Un sistema $a_1, a_2, \ldots, a_m$ de vectores de un espacio unitario se llama *ortogonal* si cualesquiera dos vectores a_j y a_k $(j \neq k)$ del mismo son ortogonales. Si el sistema contiene solamente un vector, también lo llamaremos ortogonal.

TEOREMA 1. *Todo sistema ortogonal de vectores no nulos de un espacio unitario* $\mathfrak{L}$ *es linealmente independiente.*

En efecto, sea $a_1, a_2, \ldots, a_m$ un sistema ortogonal de vectores no núlos y sea

$$\alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_m a_m = o.$$

Multiplicando esta igualdad escalarmente por a_j, obtenemos

$$\alpha_1 (a_1, a_j) + \alpha_2 (a_2, a_j) + \ldots + \alpha_m (a_m, a_j) = 0$$

o

$$\alpha_j (a_j, a_j) = 0, \tag{13}$$

ya que debido a la ortogonalidad del sistema todos los demás términos se anulan. Pero $a_j \neq o$; por consiguiente, $(a_j, a_j) \neq 0$ y de (13) resulta $\alpha_j = 0$ que es lo que se quería demostrar.

Sea n la dimensión del espacio $\mathfrak{L}$. El teorema 1 muestra que todo sistema ortogonal de vectores no nulos de $\mathfrak{L}$ no puede contener más de n vectores. Si en $\mathfrak{L}$ existen n vectores no nulos ortogonales, ellos constituyen *una base ortogonal* del espacio $\mathfrak{L}$. Probemos que en $\mathfrak{L}$ siempre existe una base de esta índole. Es más, probemos que *en todo espacio unitario* $\mathfrak{L}$ *cualquier sistema ortogonal de vectores no nulos puede ser complementado hasta obtener una base ortogonal del espacio* $\mathfrak{L}$. Sea dado en $\mathfrak{L}$ un sistema ortogonal $a_1, \ldots, a_m$. Complementémoslo hasta obtener un sistema ortogonal maximal $a_1, \ldots, a_m, a_{m+1}, \ldots, a_s$ de vectores no nulos. Tal completación es posible, ya que según el teorema 1 ningún sistema ortogonal puede contener más de n vectores diferentes de cero. Probemos que $a_1, \ldots, a_m, \ldots \ldots, a_s$ es precisamente la base requerida del espacio $\mathfrak{L}$. Consideremos un vector x cualquiera de $\mathfrak{L}$. Pongamos

$$\xi_k = \frac{(x, a_k)}{(a_k, a_k)} \quad (k = 1, 2, \ldots, s),$$

$$y = \xi_1 a_1 + \xi_2 a_2 + \ldots + \xi_s a_s.$$

Multiplicando escalarmente todos los términos de la última igualdad por a_k, obtenemos

$$(y, a_k) = (x, a_k),$$

de donde resulta

$$(x - y, a_k) = (x, a_k) - (y, a_k) = 0.$$

Por consiguiente, el vector $x - y$ es ortogonal a todos los vectores $a_1, \ldots, a_s$. El sistema $a_1, \ldots, a_s$ es, por hipótesis, un sistema

ortogonal maximal de vectores no nulos de $\mathfrak{L}$; luego, $x-y=o$, es decir,

$$x=\xi_1 a_1+\xi_2 a_2+\ldots+\xi_s a_s.$$

Por lo tanto, el sistema linealmente independiente $a_1, \ldots, a_s$ es tal que cualquier vector x se expresa linealmente en términos del mismo. Pero esto significa precisamente que $a_1, \ldots, a_s$ es una base del espacio $\mathfrak{L}$.

Un sistema ortogonal formado por vectores de longitud igual a la unidad se denomina *ortonormal*. Es obvio que normalizando vectores ortogonales no nulos obtenemos de nuevo vectores ortogonales. Por lo tanto, al normalizar los vectores de una base ortogonal del espacio, obtendremos una base ortonormal de este espacio. Hemos visto que todo sistema de vectores ortogonales no nulos puede ser complementado hasta obtener una base ortogonal del espacio. Por esto es válido el teorema siguiente.

TEOREMA 2. *Todo sistema ortonormal de vectores de un espacio $\mathfrak{L}$ puede ser complementado hasta obtener una base ortonormal de este espacio.*

Puesto que cualquier sistema que contiene sólo un vector es ortogonal, de aquí se deduce, en particular, que *en todo espacio unitario existe una base ortonormal.*

Un sistema de vectores $e_1, e_2, \ldots, e_n$ considerados en un orden determinado ha sido llamado sistema de coordenadas de un espacio $\mathfrak{L}$, si $e_1, \ldots, e_n$ forman una base del espacio $\mathfrak{L}$. Si $e_1, \ldots, e_n$ es una base ortonormal, también el sistema de coordenadas se denomina ortonormal. La diferencia que existe entre un sistema de coordenadas arbitrario y un sistema de coordenadas ortonormal es la misma que existe entre un sistema de coordenadas cartesiano oblicuo y un sistema ortogonal del espacio corriente. Efectivamente, sea $\mathfrak{R}$ el espacio corriente de vectores-segmentos provisto del producto escalar habitual. Una base arbitraria del espacio $\mathfrak{R}$ está formada por tres vectores cualesquiera a_1, a_2 y a_3 que parten de un punto O y que no pertenecen a un mismo plano (véase la fig. 1 de la pág. 79). Tomemos un punto A y expresemos el vector $\overrightarrow{OA}$ en términos de los vectores coordenados a_1, a_2 y a_3. Sea

$$\overrightarrow{OA}=\alpha_1 a_1+\alpha_2 a_2+\alpha_3 a_3.$$

Los números α_1, α_2 y α_3 son, según la definición, las coordenadas del vector $\overrightarrow{OA}$. Al mismo tiempo, se ve de la figura que estos números son las coordenadas del punto A calculadas en el sistema de coordenadas cartesiano oblicuo con los ejes a_1, a_2 y a_3, en el que los vectores a_1, a_2 y a_3 se han tomado como segmentos unidades a lo largo de los ejes.

Por consiguiente, un sistema arbitrario de coordenadas en $\mathfrak{L}$ es equivalente, desde el punto de vista de la Geometría analítica, a un sistema de coordenadas cartesiano oblicuo con distintas unidades a lo largo de los ejes coordenados. Por el contrario, un sistema ortonormal de coordenadas e_1, e_2 y e_3 será equivalente, en este sentido, al sistema de coordenadas cartesiano ortogonal habitual con unidades iguales sobre los ejes (fig. 4).

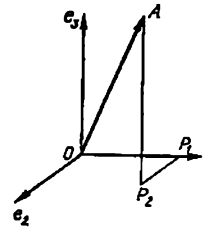


Fig. 4.

Los sistemas ortonormales de coordenadas de los espacios unitarios tienen varias propiedades específicas. Algunas de éstas serán consideradas ahora.

Si e_1, e_2, ..., e_n es un sistema ortonormal de coordenadas de un espacio $\mathfrak{L}$, las coordenadas de un vector cualquiera a son iguales respectivamente a los productos escalares (a, e_1), (a, e_2), ..., (a, e_n).

En efecto, sea

$$a = \alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_n e_n.$$

Multiplicando esta igualdad escalarmente por e_k y teniendo en cuenta que $(e_j, e_k) = 0$ para $j \neq k$, obtenemos

$$(a, e_k) = \alpha_k (e_k, e_k) = \alpha_k \quad (k = 1, 2, \ldots, n)$$

que es lo que quería demostrar.

Si en un sistema de coordenadas ortonormal los vectores a y b tienen respectivamente las coordenadas α_1, α_2, ..., α_n y β_1, β_2, ..., β_n, se tiene

$$(a, b) = \alpha_1 \overline{\beta}_1 + \alpha_2 \overline{\beta}_2 + \ldots + \alpha_n \overline{\beta}_n. \tag{14}$$

Efectivamente, sea $e_1, e_2, \ldots, e_n$ un sistema de coordenadas ortonormal dado y sea

$$a = \alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_n e_n,$$
$$b = \beta_1 e_1 + \beta_2 e_2 + \ldots + \beta_n e_n.$$

Entonces,

$$(a, b) = \left(\sum \alpha_j e_j, \quad \sum \beta_k e_k\right) = \sum\sum \alpha_j \overline{\beta}_k (e_j, e_k) = \sum \alpha_j \overline{\beta}_j.$$

DESIGUALDAD DE BESSEL. *Si $e_1, e_2, \ldots, e_m$ es un sistema ortonormal cualquiera de vectores de un espacio unitario y a es un vector arbitrario del mismo y si tomamos*

$$(a, e_k) = \alpha_k \quad (k = 1, 2, \ldots, m),$$

tiene lugar la desigualdad

$$(a, a) \geqslant |\alpha_1|^2 + |\alpha_2|^2 + \ldots + |\alpha_m|^2.$$

Daremos una demostración independiente de esta igualdad, aunque la misma se puede deducir fácilmente de las proposiciones anteriores. Consideremos un vector auxiliar

$$x = \alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_m e_m.$$

Tenemos

$$(a - x,\ a - x) = (a, a) - (a, x) - (x, a) + (x, x) \geqslant 0, \tag{15}$$

$$(x, x) = \left(\sum \alpha_j e_j,\ \sum \alpha_j e_j\right) = \sum \alpha_j \bar{\alpha}_j, \tag{16}$$

$$(a, x) = \left(a, \sum \alpha_j e_j\right) = \sum \alpha_j \bar{\alpha}_j, \tag{17}$$

$$(x, a) = \overline{(a, x)} = \sum \alpha_j \bar{\alpha}_j. \tag{18}$$

Introduciendo (16), (17) y (18) en la desigualdad (15) obtenemos

$$(a, a) - \sum \alpha_j \bar{\alpha}_j - \sum \alpha_j \bar{\alpha}_j + \sum \alpha_j \bar{\alpha}_j \geqslant 0,$$

es decir,

$$(a, a) \geqslant \sum \alpha_j \bar{\alpha}_j$$

que es lo que se quería demostrar.

Si $e_1, e_2, \ldots, e_n$ es una base ortonormal, en lugar de la desigualdad de Bessel tendremos, según (14), la igualdad

$$(a, a) = \alpha_1 \bar{\alpha}_1 + \alpha_2 \bar{\alpha}_2 + \ldots + \alpha_n \bar{\alpha}_n,$$

que se llama *igualdad de Parseval*. La igualdad de Parseval, además de ser una condición necesaria, es también una condición suficiente para que un sistema $e_1, e_2, \ldots, e_n$ ortonormal sea una base.

Si $e_1, e_2, \ldots, e_n$ es un sistema ortonormal de vectores de un espacio $\mathfrak{L}$ y si para todo vector a de $\mathfrak{L}$ se tiene

$$(a, a) = \alpha_1 \bar{\alpha}_1 + \alpha_2 \bar{\alpha}_2 + \ldots + \alpha_n \bar{\alpha}_n,$$

donde $\alpha_j = (a, e_j)$, $j = 1, 2, \ldots, n$, entonces el sistema $e_1, e_2, \ldots, e_n$ es una base del espacio $\mathfrak{L}$.

La demostración se deduce fácilmente de las proposiciones anteriores y queda a cargo del lector. •

Hemos demostrado la existencia de una base ortonormal en todo espacio unitario mediante razonamientos indirectos. No obstante, existe un método directo que permite obtener a partir de cualquier base de un espacio una base ortonormal del mismo. Este método lleva el nombre de proceso de *ortonormalización de Gram—Schmidt* y se emplea con frecuencia al considerar los espacios de funciones. Su esencia consiste en lo siguiente. Sea $a_1, \ldots, a_m$ un sistema linealmente independiente de vectores de un espacio unitario $\mathfrak{L}$. Planteémonos la tarea de construir un sistema ortonormal de vectores $e_1, e_2, \ldots, e_m$ tal que todo vector j-ésimo e_j del mismo se exprese linealmente en términos de los j primeros vectores $a_1, \ldots, a_j$. Realizaremos la construcción por inducción. El vector e_1 debe expresarse, según lo convenido, en términos de a_1 y debe ser de longitud 1. Un vector así se obtiene normalizando a_1:

$$e_1 = \frac{1}{\sqrt{(a_1, a_1)}} a_1.$$

Supongamos ahora que ya han sido construidos los vectores $e_1, \ldots, e_j$ con las propiedades requeridas correspondientes a un valor de j. Busquemos e_{j+1}. Ante todo, escojamos unos números $\alpha_1, \ldots, \alpha_j$ de modo que el vector auxiliar

$$e'_{j+1} = a_{j+1} + \alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_j e_j \tag{19}$$

sea ortogonal a los vectores $e_1, e_2, \ldots, e_j$. Multiplicando (19) escalarmente por e_k ($k = 1, 2, \ldots, j$) veremos que para ello es necesario que

$$\alpha_k = -(a_{j+1}, e_k). \tag{20}$$

Recíprocamente, tomando para α_k los valores de (20) e introduciéndolos en (19), obtenemos un vector e'_{j+1} ortogonal a $e_1, e_2, \ldots, e_j$. Puesto que a_{j+1} no puede expresarse linealmente en términos de $a_1, \ldots, a_j$ y, por consiguiente, no puede ser combinación lineal de los vectores $e_1, \ldots, e_j$, resulta que e'_{j+1} es diferente de cero y, por lo tanto, puede ser normalizado. Tomemos

$$e_{j+1} = \frac{1}{|e'_{j+1}|} e'_{j+1}. \tag{21}$$

Las relaciones (21) y (19) muestran que e_{j+1} se expresa linealmente en términos de $a_1, a_2, \ldots, a_{j+1}$. Además, e_{j+1} está normalizado y es ortogonal a todos los vectores $e_1, \ldots, e_j$. Por consiguiente, se cumplen las suposiciones de inducción y podemos considerar que la sucesión $e_1, e_2, \ldots, e_m$ ha sido construida. Si la sucesión inicial $a_1, a_2, \ldots, a_m$ era una base del espacio, es obvio que la sucesión $e_1, e_2, \ldots, e_m$ obtenida mediante el proceso de ortonormalización también será una base del espacio $\mathfrak{L}$.

17.4. Isomorfismo. En el p. 4.3 hemos quedado en llamar isomorfos dos espacios lineales si entre los elementos de los mismos se puede establecer una correspondencia biyectiva que conserva las operaciones de adición y de multiplicación por número. En los espacios unitarios a estas operaciones se agrega además la de multiplicación escalar. Por esto resulta natural llamar isomorfos los espacios unitarios sólo en el caso en el que ellos se comportan idénticamente respecto a las tres operaciones mencionadas.

DEFINICIÓN. *Dos espacios unitarios* $\mathfrak{L}$ y $\mathfrak{L}_1$ *sobre un mismo campo de coeficientes se llaman isomorfos, si entre sus elementos se puede establecer una correspondencia biyectiva en la que la suma de dos vectores de* $\mathfrak{L}$ *se transforma en la suma de los vectores correspondientes de* $\mathfrak{L}_1$, *el producto de un número por un vector de* $\mathfrak{L}$ *se transforma en el producto del mismo número por el vector correspondiente de* $\mathfrak{L}_1$ *y los productos escalares de pares correspondientes de vectores de* $\mathfrak{L}$ *y de* $\mathfrak{L}_1$ *coinciden.*

Tendrán interés para nosotros sólo aquellas propiedades de los espacios unitarios que sean propiedades de las tres operaciones principales definidas en estos espacios y que no dependan de la naturaleza de los elementos que constituyen los espacios. Desde este punto de vista los espacios unitarios isomorfos tendrán las mismas propiedades. De aquí se ve la importancia de saber clasificar, salvo un isomorfismo, todos los espacios unitarios. Esta clasificación no difiere

en nada de la clasificación de los espacios lineales y queda determinada por el teorema siguiente.

TEOREMA 3. *Para que dos espacios unitarios sobre un mismo campo de coeficientes sean isomorfos es necesario y suficiente que coincidan las dimensiones de estos espacios.*

En efecto, si dos espacios unitarios $\mathfrak{L}$ y $\mathfrak{L}_1$ son isomorfos, también serán isomorfos como espacios lineales, es decir, respecto a las operaciones de adición y de multiplicación por número. Pero los espacios lineales isomorfos son de la misma dimensión y, por consiguiente, las dimensiones de los espacios $\mathfrak{L}$ y $\mathfrak{L}_1$ coinciden. Hemos demostrado la necesidad. Recíprocamente, supongamos que las dimensiones de los espacios $\mathfrak{L}$ y $\mathfrak{L}_1$ coinciden. Tomemos en $\mathfrak{L}$ y $\mathfrak{L}_1$ unas bases ortonormales $e_1, \ldots, e_n$ y $e'_1, \ldots, e'_n$. Diremos que los vectores $x \in \mathfrak{L}$ y $x' \in \mathfrak{L}_1$ son *correspondientes* si sus coordenadas en las bases escogidas coinciden. Esta correspondencia es biyectiva y conserva las operaciones de adición y de multiplicación por número (p. 4.3). Por ello sólo debemos probar que los productos escalares de los pares correspondientes de vectores son idénticos. Consideremos dos vectores cualesquiera

$$a = \alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_n e_n,$$
$$b = \beta_1 e_1 + \beta_2 e_2 + \ldots + \beta_n e_n.$$

del espacio $\mathfrak{L}$. Los vectores correspondientes de $\mathfrak{L}_1$ son

$$a' = \alpha_1 e'_1 + \alpha_2 e'_2 + \ldots + \alpha_n e'_n,$$
$$b' = \beta_1 e'_1 + \beta_2 e'_2 + \ldots + \beta_n e'_n.$$

Puesto que las bases $e_1, \ldots, e_n$ y $e'_1, \ldots, e'_n$ son ortonormales, tenemos en virtud de la fórmula (14)

$$(a, b) = \alpha_1 \overline{\beta}_1 + \alpha_2 \overline{\beta}_2 + \ldots + \alpha_n \overline{\beta}_n = (a', b')$$

que es lo que se quería demostrar.

17.5. Sumas ortogonales. Proyecciones. Dos conjuntos de vectores $\mathfrak{M}$ y $\mathfrak{N}$ de un espacio unitario $\mathfrak{L}$ se llaman *ortogonales*, si todo vector del primer conjunto es ortogonal a todo vector del segundo. En particular, se dice que el vector a es ortogonal al conjunto $\mathfrak{M}$, si a es ortogonal a todo vector de $\mathfrak{M}$. A veces, la ortogonalidad de $\mathfrak{M}$ y $\mathfrak{N}$ se indica simbólicamente por $\mathfrak{M} \perp \mathfrak{N}$.

Si dos conjuntos $\mathfrak{M}$ y $\mathfrak{N}$ son ortogonales, la intersección de los mismos o bien es vacía, o bien consta del vector nulo únicamente.

En efecto, si el vector a está contenido en $\mathfrak{M}$ y en $\mathfrak{N}$, se tiene $(a, a) = 0$, de donde $a = o$.

Una suma $\mathfrak{A}_1 + \ldots + \mathfrak{A}_m$ de varios subespacios lineales se llama *ortogonal* si cualesquiera dos subespacios $\mathfrak{A}_j$ y $\mathfrak{A}_k$ $(j \neq k)$ son ortogonales.

Una suma ortogonal de subespacios es siempre una suma directa.

Efectivamente, si la suma es sólo de dos sumandos, la intersección de los mismos, en virtud de la observación anterior, consta solamente del vector nulo y, por consiguiente, la suma es directa. En el caso general, la demostración se realiza por inducción.

Si la suma

$$\mathfrak{A} = \mathfrak{A}_1 + \mathfrak{A}_2 + \ldots + \mathfrak{A}_s$$

es ortogonal y si

$$a = a_1 + a_2 + \ldots + a_s,$$
$$b = b_1 + b_2 + \ldots + b_s,$$

donde $a_j \in \mathfrak{A}_j$ y $b_j \in \mathfrak{A}_j$, $j = 1, \ldots, s$, *se tiene*

$$(a, b) = (a_1, b_1) + (a_2, b_2) + \ldots + (a_s, b_s). \qquad (22)$$

En efecto, puesto que $\mathfrak{A}_j \perp \mathfrak{A}_k$ para $j \neq k$, resulta que $(a_j, b_k) = 0$. Por consiguiente,

$$(a, b) = (\sum a_j, \sum b_j) = \sum\sum (a_j, b_k) = \sum (a_j, b_j)$$

que es lo que se quería demostrar.

Consideremos ahora un conjunto no vacío cualquiera $\mathfrak{M}$ de vectores de un espacio unitario $\mathfrak{L}$. El conjunto de todos los vectores del espacio $\mathfrak{L}$ ortogonales a $\mathfrak{M}$ se llama *complemento ortogonal* del conjunto $\mathfrak{M}$ y se indica por $\mathfrak{M}^\perp$.

El complemento ortogonal de un conjunto no vacío cualquiera $\mathfrak{M}$ *es un subespacio lineal.*

En efecto, si a y b pertenecen al complemento ortogonal $\mathfrak{M}^\perp$ y c es un vector cualquiera de $\mathfrak{M}$, se tiene

$$(\alpha a + \beta b, c) = \alpha (a, c) + \beta (b, c) = 0.$$

Por consiguiente, cualesquiera que sean α y β el vector $\alpha a + \beta b$ está contenido en $\mathfrak{M}^\perp$ y $\mathfrak{M}^\perp$ es un subespacio lineal.

TEOREMA 4. *Todo espacio unitario* $\mathfrak{L}$ *es la suma directa de cualquier subespacio lineal suyo* $\mathfrak{A}$ *y de su complemento ortogonal* $\mathfrak{A}^\perp$.

Sea $e_1, e_2, \ldots, e_m$ una base ortonormal del subespacio $\mathfrak{A}$ y sea $e_{m+1}, \ldots, e_s$ una base ortonormal del subespacio $\mathfrak{A}^\perp$. Para demostrar el teorema es suficiente ver que $e_1, \ldots, e_m, \ldots, e_s$ es una base del espacio $\mathfrak{L}$. Supongamos, por el contrario, que el sistema $e_1, \ldots, e_s$ no es una base del espacio $\mathfrak{L}$. Entonces, de acuerdo con el teorema 2, este sistema puede ser complementado hasta obtener una base ortonormal del espacio $\mathfrak{L}$. Sea e uno de los vectores complementarios. Puesto que e es ortogonal a todos los vectores $e_1, \ldots, e_m$, el vector e está contenido en $\mathfrak{A}^\perp$. Por consiguiente, $\mathfrak{A}^\perp$ contiene un sistema ortonormal y, por ende, linealmente independiente de vectores $e_{m+1}, \ldots, e_s, e$. Pero esto contradice a nuestra hipótesis de que $e_{m+1}, \ldots, e_s$ es una base de $\mathfrak{A}^\perp$. Hemos demostrado el teorema.

Del teorema 4 se desprende, en particular, que

$$\mathfrak{A}^{\perp\perp} = \mathfrak{A}. \tag{23}$$

Efectivamente, $\mathfrak{A}^{\perp\perp}$ está contenido indudablemente en $\mathfrak{A}$. Por otra parte, según el teorema 4, para cualquier x de $\mathfrak{A}^{\perp\perp}$ tenemos

$$x = a + b, \quad a \in \mathfrak{A} \quad \text{y} \quad b \in \mathfrak{A}^{\perp}.$$

Multiplicando escalarmente esta igualdad por b, obtenemos $(b, b) = 0$, es decir, $b = o$ y $x = a$. Luego, $\mathfrak{A}^{\perp\perp} \subset \mathfrak{A}$.

Definemos, para concluir, el concepto de proyección ortogonal de un vector sobre un subespacio lineal. Sea $\mathfrak{A}$ un subespacio lineal de un espacio unitario $\mathfrak{L}$. En virtud del teorema 4, $\mathfrak{L}$ es la suma directa del subespacio $\mathfrak{A}$ y de su complemento ortogonal $\mathfrak{A}^{\perp}$. Por consiguiente, todo vector x de $\mathfrak{L}$ puede ser representado univocamente en forma de una suma

$$x = a + b \quad (a \in \mathfrak{A} \quad \text{y} \quad b \in \mathfrak{A}^{\perp}). \tag{24}$$

El sumando a se llama *proyección del vector x sobre el subespacio $\mathfrak{A}$*. Puesto que de acuerdo con la fórmula (23) $\mathfrak{A}$ es el complemento ortogonal de $\mathfrak{A}^{\perp}$, el sumando b de la igualdad (24) representa la proyección del vector x sobre el subespacio $\mathfrak{A}^{\perp}$.

Multiplicando (24) por un número α, obtenemos

$$\alpha x = \alpha a + \alpha b \quad (\alpha a \in \mathfrak{A} \quad \text{y} \quad \alpha b \in \mathfrak{A}^{\perp}),$$

es decir, *la proyección del producto de un número por un vector es igual al producto de este número por la proyección del vector.*

Análogamente se demuestra también la proposición de que *la proyección de una suma de vectores sobre un subespacio es igual a la suma de las proyecciones de los sumandos sobre este subespacio.*

Ejemplos y problemas.

1. Sea e_1, e_2 y e_3 un sistema ortonormal de coordenadas de un espacio unitario de tres dimensiones. Pruébese que el sistema de vectores

$$a_1 = \frac{1}{3}(2e_1 + 2e_2 - e_3),$$

$$a_2 = \frac{1}{3}(2e_1 - e_2 + 2e_3),$$

$$a_3 = \frac{1}{3}(e_1 - 2e_2 + 2e_3)$$

también constituye un sistema ortonormal de coordenadas de este espacio.

2. En un espacio unitario $\mathfrak{L}$ se ha tomado un sistema ortonormal de coordenadas $e_1, \ldots, e_n$. Pruébese que un sistema de vectores $a_1, \ldots, a_n$ será también una base ortonormal del espacio $\mathfrak{L}$ cuando, y sólo cuando, la matriz formada por las filas coordenadas de estos vectores sea unitaria (véase el p. 1.3).

3. Si $a_1, \ldots, a_m$ es una base ortonormal de un subespacio lineal $\mathfrak{A}$, la proyección de un vector x sobre $\mathfrak{A}$ es igual a

$$(x, a_1)\, a_1 + (x, a_2)\, a_2 + \ldots + (x, a_m)\, a_m.$$

4. En el espacio de todos los polinomios en λ de grado no mayor que n el producto escalar se define mediante la fórmula (7) del p. 17.1. Pruébese que éste será un espacio unitario de dimensión $n+1$. Pruébese que ortonormalizando según Gram—Schmidt en este espacio la sucesión 1, λ y λ^2 obtenemos los polinomios 1, $\sqrt{3}\,(2\lambda-1)$ y $\sqrt{5}\,(6\lambda^2-6\lambda+1)$.

5. Se llama *determinante de Gram* de un sistema de vectores $a_1, a_2, \ldots, a_n$ de un espacio unitario $\mathfrak{L}$ de n dimensiones el determinante

$$\Delta = \begin{vmatrix} (a_1, a_1) & (a_1, a_2) & \ldots & (a_1, a_n) \\ (a_2, a_1) & (a_2, a_2) & \ldots & (a_2, a_n) \\ \ldots & \ldots & \ldots & \ldots \\ (a_n, a_1) & (a_n, a_2) & \ldots & (a_n, a_n) \end{vmatrix}.$$

Supongamos que en $\mathfrak{L}$ se ha escogido un sistema ortonormal de coordenadas. Pruébese que el determinante de Gram Δ es igual al cuadrado del módulo del determinante formado por las filas coordenadas de los vectores $a_1, \ldots, a_n$. Pruébese también que $a_1, a_2, \ldots, a_n$ son linealmente independientes cuando, y sólo cuando, el determinante de Gram de los mismos sea diferente de cero.

6. Dése una interpretación geométrica al proceso de ortonormalización de Gram—Schmidt en el caso del espacio habitual de tres dimensiones de los vectores-segmentos.

§ 18. Aplicaciones conjugadas

18. 1. Funciones lineales. Sea $\mathfrak{L}$ un espacio lineal cualquiera de dimensión finita sobre un campo K. Pongamos en correspondencia a todo vector x de $\mathfrak{L}$ un número $f(x)$ de K. Las correspondencias de este tipo han sido llamadas en el p. 4.1 funciones con valores en K definidas sobre $\mathfrak{L}$. Una función $f(x)$ se llama *lineal* si para cualesquiera x e y de $\mathfrak{L}$ y cualesquiera α y β de K se tiene

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y). \qquad (1)$$

Tomando en (1) $\alpha = \beta = 0$, obtenemos

$$f(o) = 0.$$

Por esta razón en lugar de «función lineal» se dice a veces «función lineal *homogénea*».

Es fácil ver que *la suma de funciones lineales y el producto de un número por una función lineal son de nuevo funciones lineales.* Probemos, por ejemplo, la primera proposición. Sea $f = g + h$, donde g y h son funciones lineales. De acuerdo con la definición

$$f(\alpha x + \beta y) = g(\alpha x + \beta y) + h(\alpha x + \beta y) =$$
$$= \alpha g(x) + \beta g(y) + \alpha h(x) + \beta h(y) = \alpha f(x) + \beta f(y)$$

que es lo que se quería demostrar.

Conocemos ya que las operaciones de adición de funciones y de multiplicación de las mismas por números satisfacen los axiomas de un espacio lineal. Puesto que la suma de funciones lineales y el producto de una función lineal por un número son de nuevo funciones lineales, resulta que el conjunto de todas las funciones lineales, definidas sobre un espacio vectorial $\mathfrak{L}$, es por sí mismo un espacio lineal. Este espacio se llama espacio *conjugado* (o dual) respecto a $\mathfrak{L}$ y se indica por $\mathfrak{L}'$.

De la igualdad (1) que caracteriza las funciones lineales se desprende inmediatamente una relación más general

$$f(\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_m x_m) = \alpha_1 f(x_1) + \alpha_2 f(x_2) + \ldots + \alpha_m f(x_m), \quad (2)$$

cuya demostración omitimos ya que es obvia. Los teoremas que siguen esclarecen en gran medida la estructura de las funciones lineales.

TEOREMA 1. *Sea $e_1, e_2, \ldots, e_n$ un sistema de coordenadas cualquiera de un espacio lineal $\mathfrak{L}$. Tomemos una sucesión totalmente arbitraria $\alpha_1, \alpha_2, \ldots, \alpha_n$ de números de K. Entonces existe una función lineal $f(x)$, y sólo una, que está definida sobre $\mathfrak{L}$ y que satisface las condiciones*

$$f(e_j) = \alpha_j \quad (j = 1, 2, \ldots, n). \quad (3)$$

DEMOSTRACIÓN. Sea x un vector de $\mathfrak{L}$:

$$x = \xi_1 e_1 + \xi_2 e_2 + \ldots + \xi_n e_n. \quad (4)$$

Poniendo en correspondencia a este vector el número $\alpha_1\xi_1 + \alpha_2\xi_2 + \ldots + \alpha_n\xi_n$, obtenemos una función $f(x)$ sobre $\mathfrak{L}$. Es decir, por definición

$$f(x) = \alpha_1\xi_1 + \alpha_2\xi_2 + \ldots + \alpha_n\xi_n.$$

Tomando aquí $x = e_j$ tendremos $f(e_j) = \alpha_j$ $(j = 1, \ldots, n)$. Por otra parte, si

$$y = \eta_1 e_1 + \eta_2 e_2 + \ldots + \eta_n e_n, \quad (5)$$

tenemos

$$f(y) = \alpha_1\eta_1 + \alpha_2\eta_2 + \ldots + \alpha_n\eta_n.$$

De (4) y de (5) se deduce que

$$\alpha x + \beta y = (\alpha\xi_1 + \beta\eta_1) e_1 + \ldots + (\alpha\xi_n + \beta\eta_n) e_n,$$

de donde

$$f(\alpha x + \beta y) = \alpha_1 (\alpha\xi_1 + \beta\eta_1) + \ldots + \alpha_n (\alpha\xi_n + \beta\eta_n) = \alpha f(x) + \beta f(y).$$

Por consiguiente, $f(x)$ es una función lineal.

Hemos demostrado que existe una función lineal que satisface las condiciones (3). Probemos que es única. Sea $g(x)$ una función lineal tal que $g(e_j) = \alpha_j$ $(j = 1, \ldots, n)$. Entonces para un vector

cualquiera x de coordenadas $\xi_1, \ldots, \xi_n$ tendremos

$$g(x) = g(\xi_1 e_1 + \ldots + \xi_n e_n) = \xi_1 g(e_1) + \ldots + \xi_n g(e_n) = \\ = \xi_1 \alpha_1 + \ldots + \xi_n \alpha_n,$$

es decir, $g(x) = f(x)$.

El teorema 1 ha sido demostrado para espacios lineales $\mathfrak{L}$ cualesquiera. Si el espacio $\mathfrak{L}$ es unitario, las funciones lineales sobre $\mathfrak{L}$ tienen una expresión muy sencilla en términos del producto escalar.

TEOREMA 2. *El producto escalar (x, a) de vectores de un espacio unitario $\mathfrak{L}$ es una función lineal de x siendo a fijo. Con ello a todo vector a de $\mathfrak{L}$ se pone en correspondencia una función lineal definida sobre $\mathfrak{L}$. Además, para distintos vectores se obtienen distintas funciones lineales y todas las funciones lineales sobre $\mathfrak{L}$ pueden ser construidas por este procedimiento.*

La primera afirmación es evidente, ya que de $f(x) = (x, a)$ se deduce que

$$f(\alpha x + \beta y) = (\alpha x + \beta y, a) = \alpha (x, a) + \beta (y, a) = \alpha f(x) + \beta f(y).$$

Demostremos la segunda afirmación. Supongamos que, al contrario, a diferentes vectores a y b corresponde una misma función lineal; entonces esto significa que la igualdad $(x, a) = (x, b)$ tiene lugar para cualesquiera vectores x de $\mathfrak{L}$. Pasando los términos a un mismo miembro, podemos representarla en la forma $(x, a-b) = 0$, de donde tomando $x = a-b$ obtenemos $(a-b, a-b) = 0$. Pero el vector nulo es el único vector cuyo cuadrado escalar es igual a cero, es decir $a-b = o$ o $a = b$. Resta demostrar la tercera afirmación, esto es, que toda función lineal $f(x)$ definida sobre $\mathfrak{L}$ puede ser representada en la forma

$$f(x) = (x, a), \tag{6}$$

donde a depende de f y x es un vector arbitrario de $\mathfrak{L}$. Indiquemos por $\mathfrak{A}$ el conjunto de todos los vectores para los cuales $f(x) = 0$. Puesto que $f(o) = 0$, el vector o indudablemente pertenece a $\mathfrak{A}$. Además, si los vectores a y b pertenecen a $\mathfrak{A}$, se tiene

$$f(\alpha a + \beta b) = \alpha f(a) + \beta f(b) = 0,$$

es decir, también $\alpha a + \beta b$ pertenece a $\mathfrak{A}$. Por consiguiente, $\mathfrak{A}$ es un subespacio lineal del espacio $\mathfrak{L}$. Si $\mathfrak{A} = \mathfrak{L}$, resulta que $f(x) = 0$ para todos los vectores x. Para satisfacer la condición (6) es suficiente tomar en este caso $a = o$. Supongamos por ello que $\mathfrak{A} \neq \mathfrak{L}$. Tomemos un vector no nulo b cualquiera de $\mathfrak{L}$ que sea ortogonal a $\mathfrak{A}$ y tratemos de determinar un número α tal que el vector $a = \alpha b$ satisfaga la relación (6). Sea $f(b) = \beta$ y sea $f(x) = \xi$, donde x es un vector cualquiera de $\mathfrak{L}$. Tenemos

$$f\left(x - \frac{\xi}{\beta} b\right) = f(x) - \frac{\xi}{\beta} f(b) = 0.$$

Indicando por c la diferencia $x-\frac{\xi}{\beta}b$, veremos que c pertenece a $\mathfrak{A}$ y que $x=c+\frac{\xi}{\beta}b$. De aquí resulta

$$(x, a)=\left(\frac{\xi}{\beta}b+c, \alpha b\right)=\frac{\bar{\alpha}\xi}{\beta}(b, b)+\bar{\alpha}(c, b).$$

Puesto que el vector b es ortogonal a $\mathfrak{A}$, tenemos $(c, b)=0$ y, por consiguiente,

$$(x, a)=\frac{\bar{\alpha}\xi}{\beta}(b, b). \tag{7}$$

La igualdad (7) muestra que siendo $\alpha=\bar{\beta}(b, b)^{-1}$ se tiene para cualquier x

$$(x, a)=\xi=f(x)$$

que es lo que se quería demostrar.

18.2. Aplicaciones conjugadas. Emplearemos ahora los resultados del punto anterior con el fin de obtener para toda aplicación lineal de un espacio unitario una aplicación nueva univocamente determinada y llamada *conjugada* de la dada.

Sea $\mathcal{A}$ una aplicación lineal de un espacio unitario $\mathfrak{L}$. Tomemos en $\mathfrak{L}$ un vector arbitrario y y consideremos la expresión

$$f(x)=(x\mathcal{A}, y), \tag{8}$$

donde x es un vector variable. Puesto que

$$f(\alpha u+\beta v)+((\alpha u+\beta v)\mathcal{A}, y)=(\alpha\cdot u\mathcal{A}+\beta\cdot v\mathcal{A}, y)=\alpha f(u)+\beta f(v),$$

$f(x)$ es una función lineal. Por esto $f(x)$ puede ser representada, de acuerdo con el teorema 2, en la forma

$$f(x)=(x, a), \tag{9}$$

donde el vector a queda determinado univocamente por la función $f(x)$, es decir, por la aplicación $\mathcal{A}$ y por el vector y. Si consideramos $\mathcal{A}$ como una aplicación dada y hacemos variar el vector y, tendremos entonces para todo vector y un vector determinado a. La aplicación que transforma y en a se indica por $\mathcal{A}^*$ y se llama *conjugada* de $\mathcal{A}$; es decir, $a=y\mathcal{A}^*$. Introduciendo este resultado en (9) y comparando con (8) obtenemos la relación

$$(x\mathcal{A}, y)=(x, y\mathcal{A}^*) \tag{10}$$

que tiene lugar para cualesquiera vectores x e y de $\mathfrak{L}$.

La propiedad (10) caracteriza plenamente la aplicación conjugada $\mathcal{A}^*$. En efecto, sea $\mathcal{B}$ una aplicación con la misma propiedad, es decir, que para cualesquiera x e y

$$(x\mathcal{A}, y)=(x, y\mathcal{B})$$

o

$$(x, y\mathcal{A}^*) - (x, y\mathcal{B}) = (x, y\mathcal{A}^* - y\mathcal{B}) = 0.$$

Esto significa que el vector $y\mathcal{A}^* - y\mathcal{B}$ es ortogonal a todo el espacio, de donde resulta que

$$y\mathcal{A}^* = y\mathcal{B}.$$

La última igualdad tiene lugar para todos los y y, por consiguiente, $\mathcal{A}^* = \mathcal{B}$.

Es fácil demostrar ahora que *la aplicación conjugada $\mathcal{A}^*$ es lineal*. Efectivamente, en virtud de (10) se tiene

$$(x, (\alpha u + \beta v)\mathcal{A}^*) = (x\mathcal{A}, \alpha u + \beta v) = \bar{\alpha}(x\mathcal{A}, u) + \bar{\beta}(x\mathcal{A}, v) =$$
$$= \bar{\alpha}(x, u\mathcal{A}^*) + \bar{\beta}(x, v\mathcal{A}^*) = (x, \alpha \cdot u\mathcal{A}^* + \beta \cdot v\mathcal{A}^*)$$

cualquiera que sea el vector x. Debido a la segunda afirmación del teorema 2 esto implica la igualdad

$$(\alpha u + \beta v)\mathcal{A}^* = \alpha \cdot u\mathcal{A}^* + \beta \cdot v\mathcal{A}^*$$

que significa precisamente que $\mathcal{A}^*$ es lineal.

La operación de paso a la aplicación conjugada posee las propiedades siguientes:

a) $(\mathcal{A}^*)^* = \mathcal{A}$,
b) $(\alpha\mathcal{A})^* = \bar{\alpha}\mathcal{A}^*$,
c) $(\mathcal{A} + \mathcal{B})^* = \mathcal{A}^* + \mathcal{B}^*$,
d) $(\mathcal{A}\mathcal{B})^* = \mathcal{B}^*\mathcal{A}^*$.

Todas estas propiedades se demuestran de un mismo modo y por ello nos limitaremos a demostrar sólo una de ellas.

Por ejemplo,

$$(x\mathcal{A}\mathcal{B}, y) = (x\mathcal{A}, y\mathcal{B}^*) = (x, y\mathcal{B}^*\mathcal{A}^*),$$

es decir, $(\mathcal{A}\mathcal{B})^* = \mathcal{B}^*\mathcal{A}^*$.

Notemos aquí mismo que las aplicaciones conjugadas de las aplicaciones identidad y nula coinciden con ellas mismas. Efectivamente,

$$(x\mathcal{E}, y) = (x, y) = (x, y\mathcal{E}) \quad \text{y} \quad (x\mathcal{O}, y) = 0 = (x, y\mathcal{O}).$$

Veamos la relación que existe entre las matrices de las aplicaciones conjugadas. Tomemos en el espacio $\mathfrak{L}$ un sistema *ortonormal* de coordenadas $e_1, e_2, \ldots, e_n$ y sea

$$e_i\mathcal{A} = \alpha_{i1}e_1 + \alpha_{i2}e_2 + \ldots + \alpha_{in}e_n,$$
$$e_i\mathcal{A}^* = \beta_{i1}e_1 + \beta_{i2}e_2 + \ldots + \beta_{in}e_n \ (i = 1, \ldots, n).$$

Multiplicando estas igualdades escalarmente por e_j y teniendo en cuenta que el sistema $e_1, \ldots, e_n$ es ortogonal, obtenemos

$$(e_i\mathcal{A}, e_j) = \alpha_{ij} \quad \text{y} \quad (e_i\mathcal{A}^*, e_j) = \beta_{ij} \quad (i, j = 1, \ldots, n).$$

De aquí resulta que

$$\alpha_{ij} = (e_i\mathcal{A}, e_j) = (e_i, e_j\mathcal{A}^*) = \overline{(e_j\mathcal{A}^*, e_i)} = \overline{\beta}_{ji}.$$

Por consiguiente, si A es la matriz de la aplicación $\mathcal{A}$, la matriz de la aplicación conjugada $\mathcal{A}^*$ será igual a $\overline{A}'$. Hemos obtenido el siguiente teorema.

TEOREMA 3. *Si una aplicación lineal $\mathcal{A}$ tiene en un sistema ortonormal de coordenadas la matriz A, la aplicación conjugada $\mathcal{A}^*$ tendrá en este mismo sistema la matriz conjugada transpuesta $\overline{A}'$.*

La operación del paso a las aplicaciones conjugadas posee las propiedades a), b), c) y d). En base al teorema 3 deducimos de aquí que estas mismas propiedades posee también la operación del paso a las matrices conjugadas transpuestas. Este resultado se obtiene también mediante cálculo directo (compárese con el p. 1.3).

18.3. Aplicaciones normales. Una serie de propiedades notables pueden ser obtenidas en el caso de las aplicaciones lineales de un espacio unitario que conmutan con sus conjugadas, es decir, que satisfacen la relación

$$\mathcal{A}\mathcal{A}^* = \mathcal{A}^*\mathcal{A}.$$

Estas aplicaciones se llaman *normales*. Recordando que en un sistema ortonormal de coordenadas las aplicaciones conjugadas tienen matrices conjugadas transpuestas, llegamos directamente a la conclusión de que son normales aquellas aplicaciones lineales de un espacio unitario, y sólo aquéllas, cuyas matrices, calculadas en unas bases ortonormales, satisfacen la relación

$$A\overline{A}' = \overline{A}'A.$$

Con la misma facilidad se obtienen también las siguientes propiedades de las aplicaciones normales.

TEOREMA 4. *Todo vector propio a de una aplicación normal $\mathcal{A}$ correspondiente a un valor propio ρ es al mismo tiempo un vector propio de la aplicación conjugada $\mathcal{A}^*$ pero correspondiente al valor propio conjugado de ρ.*

Tenemos

$$\mathcal{A}\mathcal{A}^* = \mathcal{A}^*\mathcal{A} \quad \text{y} \quad a(\mathcal{A} - \rho\mathcal{E}) = o.$$

De aquí resulta

$$o = (a(\mathcal{A} - \rho\mathcal{E}), a(\mathcal{A} - \rho\mathcal{E})) = (a(\mathcal{A} - \rho\mathcal{E})(\mathcal{A}^* - \overline{\rho}\mathcal{E}), a) =$$
$$= (a(\mathcal{A}^* - \overline{\rho}\mathcal{E})(\mathcal{A} - \rho\mathcal{E}), a) = (a(\mathcal{A}^* - \overline{\rho}\mathcal{E}), a(\mathcal{A}^* - \overline{\rho}\mathcal{E})),$$

es decir

$$a(\mathcal{A}^* - \bar{\rho}\mathcal{E}) = o$$

que es lo que se quería demostrar.

TEOREMA 5. *Los vectores propios correspondientes a diferentes valores propios de una aplicación normal $\mathcal{A}$ son ortogonales.*

Sea

$$a\mathcal{A} = \rho a \quad \text{y} \quad b\mathcal{A} = \sigma b \quad (\rho \neq \sigma).$$

Entonces

$$\rho(a, b) = (a\mathcal{A}, b) = (a, b\mathcal{A}^*) = \sigma(a, b),$$

es decir,

$$(\rho - \sigma)(a, b) = 0 \quad \text{y} \quad (a, b) = 0.$$

Algo más compleja es la demostración del siguiente teorema principal.

TEOREMA 6. *Para toda aplicación normal $\mathcal{A}$ de un espacio unitario complejo existe una base ortonormal formada por los vectores propios de la aplicación $\mathcal{A}$; la matriz de $\mathcal{A}$ es de forma diagonal en esta base.*

Para la demostración tomamos en el espacio inicial $\mathfrak{L}$ un vector propio cualquiera $a_1 \neq o$ de la aplicación $\mathcal{A}$ e indicamos por $\mathfrak{L}_1$ el subespacio ortogonal a a_1. Si

$$a_1\mathcal{A} = \rho_1 a_1 \quad \text{y} \quad x \in \mathfrak{L}_1,$$

se tiene

$$(a_1, x\mathcal{A}) = (a_1\mathcal{A}^*, x) = \bar{\rho}_1 (a_1, x) = 0,$$

es decir, el subespacio $\mathfrak{L}_1$ es invariante respecto de $\mathcal{A}$. De la invariancia de $\mathfrak{L}_1$ se deduce que en él existe un vector propio a_2 de la aplicación $\mathcal{A}$. Indiquemos por $\mathfrak{L}_2$ el subespacio formado por todos los vectores de $\mathfrak{L}$ ortogonales a a_2 y pongamos $\mathfrak{L}_2' = \mathfrak{L}_1 \cap \mathfrak{L}_2$. Puesto que $\mathfrak{L}_1$ y $\mathfrak{L}_2$ son invariantes respecto de $\mathcal{A}$, también será invariante el espacio $\mathfrak{L}_2'$, en el cual debe existir, por consiguiente, un vector propio no nulo a_3 de la aplicación $\mathcal{A}$. Indicando por $\mathfrak{L}_3$ el conjunto de todos los vectores de $\mathfrak{L}$ ortogonales a a_3 y tomando $\mathfrak{L}_3' = \mathfrak{L}_1 \cap \mathfrak{L}_2 \cap \mathfrak{L}_3$, obtenemos un subespacio invariante de vectores ortogonales a a_1, a_2 y a_3. Continuando el proceso encontraremos la base ortogonal requerida $a_1, a_2, \ldots, a_n$ del espacio $\mathfrak{L}$ formada por los vectores propios de la aplicación $\mathcal{A}$.

La propiedad de las aplicaciones normales de los espacios complejos establecida en el teorema 6 es característica para estas aplicaciones. Efectivamente, si la matriz A de una aplicación $\mathcal{A}$ es diagonal en una base ortonormal, la matriz $\overline{A}'$ de la aplicación conjugada también será diagonal y, por consiguiente, conmutará con A.

En los espacios unitarios reales la situación es algo diferente. Para analizarla demostremos primero una proposición general relacionada con unas aplicaciones reales lineales cualesquiera.

TEOREMA 7. *Toda aplicación lineal $\mathcal{A}$ de un espacio real no nulo tiene por lo menos un subespacio invariante de dimensión 1 o 2.*

Si el polinomio característico $\varphi(\lambda)$ de la aplicación $\mathcal{A}$ tiene una raíz real α, la aplicación $\mathcal{A}$ tiene en $\mathfrak{L}$ un vector propio no nulo. El subespacio tendido sobre este vector será el subespacio invariante requerido de una dimensión.

Supongamos ahora que $\varphi(\lambda)$ no tiene raíces reales. En este caso $\varphi(\lambda)$ tendrá un par de raíces conjugadas $\alpha=\rho+i\sigma$ y $\bar{\alpha}=\rho-i\sigma$, ya que los coeficientes del polinomio $\varphi(\lambda)$ son reales. Tomemos en $\mathfrak{L}$ un sistema de coordenadas $e_1, e_2, \ldots, e_n$ y sea A la matriz de la aplicación $\mathcal{A}$ en este sistema. Consideremos la ecuación

$$[\xi_1, \xi_2, \ldots, \xi_n]\, A = \alpha\, [\xi_1, \xi_2, \ldots, \xi_n], \tag{11}$$

donde $\xi_1, \xi_2, \ldots, \xi_n$ son unas incógnitas cuyos valores determinaremos en el cuerpo de los números complejos. La ecuación (11) puede ser representada en la forma

$$[\xi_1, \xi_2, \ldots, \xi_n]\,(\alpha E - A) = O,$$

donde E es la matriz unidad y O es la fila nula. Esta ecuación equivale a un sistema de ecuaciones lineales homogéneas respecto a las incógnitas $\xi_1, \ldots, \xi_n$ de matriz $(\alpha E - A)'$ (compárese con el p. 11.4). Puesto que el determinante de esta matriz es $\varphi(\alpha)$ y, por consiguiente, es igual a cero, la ecuación (11) tiene en el cuerpo de los números complejos una solución no nula que indicaremos por las mismas letras $\xi_1, \ldots, \xi_n$. Tomemos para abreviar

$$[\xi_1, \xi_2, \ldots, \xi_n] = x,$$

de modo que la relación (11) se convierte en

$$xA = \alpha x. \tag{12}$$

Pasando aquí a los números complejos conjugados, obtenemos $\bar{x}\,\bar{A} = \bar{\alpha}\,\bar{x}$. Pero los elementos de la matriz A son reales y, por lo tanto, $\bar{A} = A$ y

$$\bar{x}A = \bar{\alpha}\,\bar{x}. \tag{13}$$

Puesto que las filas $x+\bar{x}$ e $i(x-\bar{x})$ son reales, en el espacio $\mathfrak{L}$ existen unos vectores a y b, cuyas filas coordenadas serán iguales respectivamente a

$$\left.\begin{aligned} [a] &= \tfrac{1}{2}(x+\bar{x}), \\ [b] &= \tfrac{1}{2i}(x-\bar{x}). \end{aligned}\right\} \tag{14}$$

Expresando aquí x y $\bar{x}$ en términos de $[a]$ y $[b]$ y empleando (12) y (13), llegamos a las relaciones

$$\begin{matrix}[a]\,A=\rho\,[a]-\sigma\,[b],\\ [b]\,A=\sigma\,[a]+\rho\,[b]\end{matrix}\quad(\alpha=\rho+i\sigma).$$

Estas relaciones equivalen a las igualdades

$$\left.\begin{matrix}a\mathcal{A}=\rho a-\sigma b,\\ b\mathcal{A}=\sigma a+\rho b\end{matrix}\right\} \tag{15}$$

que muestran que el subespacio tendido sobre los vectores a y b es invariante respecto a la aplicación $\mathcal{A}$. Hemos demostrado el teorema.

Supongamos ahora que el espacio considerado $\mathfrak{L}$ es un espacio unitario real y que $\mathcal{A}$ es una aplicación normal del mismo, cuyo polinomio, característico $\varphi(\lambda)$ tiene dos raíces conjugadas $\rho+i\sigma$ y $\rho-i\sigma$ $(\sigma\neq 0)$. Tomando para el sistema de coordenadas $e_1, \ldots, e_n$ de $\mathfrak{L}$ una base *ortonormal* cualquiera y repitiendo el razonamiento anterior, obtendremos de nuevo en $\mathfrak{L}$ unos vectores a y b ligados por las relaciones (15). Probemos que los vectores a y b serán ahora *ortogonales*. En efecto, para determinar estos vectores hemos tenido que considerar el espacio de filas $\tilde{\mathfrak{L}}$ sobre el cuerpo de los números complejos. Podemos aceptar que este espacio de filas es unitario, tomando, de acuerdo con el p. 17.1, para el producto escalar de las filas $[\xi_1, \ldots, \xi_n]$ y $[\eta_1, \ldots, \eta_n]$ la expresión $\xi_1\bar{\eta}_1+\ldots$ $\ldots+\xi_n\bar{\eta}_n$. Las filas $[e_1], \ldots, [e_n]$ forman una base ortonormal de $\tilde{\mathfrak{L}}$. La aplicación consistente en la multiplicación de las filas por una matriz A será una aplicación lineal y su matriz coincidirá con A en la base señalada. Puesto que $\bar{A}=A$, la aplicación considerada será normal y las filas x y $\bar{x}$, determinadas durante la demostración del teorema 7, serán unos vectores propios correspondientes a diferentes valores propios $\rho+i\sigma$ y $\rho-i\sigma$. En virtud del teorema 5 tenemos $(x, \bar{x})=0$, de donde

$$([a], [b])=\frac{1}{4i}(x+\bar{x},\ x-\bar{x})=0,$$

es decir, $(a, b)=0$. Tomando para x un vector de longitud $\sqrt{2}$, obtendremos de (14) que a y b serán de longitud 1.

Demostremos además que el subespacio de $\mathfrak{L}$ ortogonal a a y b será invariante respecto de $\mathcal{A}$. En el espacio de filas $\tilde{\mathfrak{L}}$ el subespacio ortogonal a $[a]$ y $[b]$ coincide con el subespacio $\tilde{\mathfrak{L}}_1$ ortogonal a x y $\bar{x}$. Este último subespacio es invariante respecto de $\mathcal{A}$, ya que es la intersección de subespacios ortogonales a los vectores propios x y $\bar{x}$ de una aplicación normal. Sea ahora $y\in\mathfrak{L}$ y $(a, y)=(b, y)=0$;

entonces $[y] \in \mathfrak{L}_1$ e $[y]\, A \in \mathfrak{L}_1$, de donde resulta

$$(y\mathcal{A},\ a) = ([y]\, A,\ [a]) = 0 \quad \text{y} \quad (y\mathcal{A},\ b) = 0$$

que es lo que se quería demostrar.

La aplicación $\mathcal{A}_1$ inducida por la aplicación $\mathcal{A}$ en el espacio de dos dimensiones tendido sobre los vectores a y b tiene, en virtud de (15), la matriz

$$A_1 = \begin{bmatrix} \rho & -\sigma \\ \sigma & \rho \end{bmatrix}.$$

Tomando para el número complejo $\rho + i\sigma$ la expresión $r\,(\cos\varphi - i \operatorname{sen}\varphi)$, podemos representar la matriz A_1 en la forma

$$A_1 = r \begin{bmatrix} \cos\varphi & \operatorname{sen}\varphi \\ -\operatorname{sen}\varphi & \cos\varphi \end{bmatrix}.$$

Por consiguiente, $\mathcal{A}_1$ es el producto de la aplicación de matriz rE y de la aplicación de matriz $\begin{bmatrix} \cos\varphi & \operatorname{sen}\varphi \\ -\operatorname{sen}\varphi & \cos\varphi \end{bmatrix}$. La primera es una aplicación de semejanza de centro en el origen de coordenadas y de coeficiente de dilatación igual a r y la segunda representa, como puede verse fácilmente, un giro de ángulo φ de los vectores alrededor del origen de coordenadas en el plano determinado por a y b.

TEOREMA 8. *Para toda aplicación normal $\mathcal{A}$ de un espacio real unitario $\mathfrak{L}$ existe en $\mathfrak{L}$ una base ortonormal en la que la matriz de la aplicación $\mathcal{A}$ tiene la forma*

$$A = \begin{bmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_k \end{bmatrix} + r_1 \begin{bmatrix} \cos\varphi_1 & \operatorname{sen}\varphi_1 \\ -\operatorname{sen}\varphi_1 & \cos\varphi_1 \end{bmatrix} + \dots$$

$$\dots + r_m \begin{bmatrix} \cos\varphi_m & \operatorname{sen}\varphi_m \\ -\operatorname{sen}\varphi_m & \cos\varphi_m \end{bmatrix}, \tag{16}$$

donde los números k y m pueden ser iguales a cero.

La demostración es casi idéntica a la demsotración del teorema análogo 6. La diferencia estriba sólo en que no podemos afirmar ahora que todo subespacio $\mathfrak{L}'_l$ invariante respecto de $\mathcal{A}$ contiene un vector propio no nulo a_{l+1} de la aplicación $\mathcal{A}$. Pero si $\mathfrak{L}'_l$ no contiene vectores propios de la aplicación $\mathcal{A}$, en $\mathfrak{L}'_l$ existe, en virtud del teorema 7, un par de vectores recíprocamente ortogonales a_{l+1} y b_{l+1} ligados por relaciones de tipo (15). Para el subespacio $\mathfrak{L}_{l+1}$ se puede tomar entonces el conjunto de vectores de $\mathfrak{L}$ ortogonales a a_{l+1} y b_{l+1}. De las observaciones hechas anteriormente resulta que $\mathfrak{L}_{l+1}$ será invariante respecto de $\mathcal{A}$ y el proceso se puede continuar después según el esquema expuesto en la demostración del teorema 6. Así obtendremos una base ortonormal de $\mathfrak{L}$ compuesta

por los vectores $a_1, \ldots, a_k, a_{k+1}, b_{k+1}, \ldots, a_{k+m}, b_{k+m}$ ligados por relaciones de tipo

$$a_p\mathcal{A} = \alpha_p a_p \quad (p = 1, \ldots, k),$$
$$\left.\begin{aligned} a_q\mathcal{A} &= \rho_q a_q - \sigma_q b_q \\ b_q\mathcal{A} &= \sigma_q a_q + \rho_q b_q \end{aligned}\right\} \quad (q = 1, \ldots, m, \ \sigma_q \neq 0);$$

estas últimas muestran que la matriz de la aplicación $\mathcal{A}$ tendrá en la base señalada precisamente la forma (16).

Ejemplos y problemas

1. Tomemos en el espacio euclídeo corriente de tres dimensiones $\mathfrak{R}$ una recta orientada que pasa por el origen de coordenadas e indiquemos por $f(x)$ la longitud de la proyección del vector x sobre esta recta tomada con el signo correspondiente. Pruébese que $f(x)$ es una función lineal y que toda función lineal sobre el espacio $\mathfrak{R}$ es de la forma $\alpha f(x)$, donde $\alpha \geqslant 0$ y el eje de proyección se ha escogido convenientemente (compárese con el teorema 2).

2. Una función $f(x)$ definida sobre un espacio lineal complejo se llama *antilineal*, si $f(x+y) = f(x) + f(y)$ y $f(\alpha x) = \overline{\alpha} f(x)$. Pruébese que toda función antilineal sobre un espacio unitario es de la forma (a, x).

3. Demuéstrese que la correspondencia establecida en el problema anterior entre los vectores de un espacio unitario $\mathfrak{L}$ y las funciones antilineales es un isomorfismo entre $\mathfrak{L}$ y el espacio de todas las funciones antilineales sobre $\mathfrak{L}$.

4. En un espacio unitario $\mathfrak{L}$ con una base no ortonormal a_1, a_2 y a_3 están dadas dos aplicaciones lineales $\mathcal{A}$ y $\mathcal{B}$ de matrices

$$A = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 2 & 0 \\ 0 & 2 & -1 \end{bmatrix} \quad \text{y} \quad B = \begin{bmatrix} 1 & 2 & -2 \\ 2 & -1 & 0 \\ -2 & 1 & 0 \end{bmatrix}.$$

Hállese una base *ortonormal* de $\mathfrak{L}$, si se sabe que $\mathcal{A}$ y $\mathcal{B}$ son normales y que el vector a_1 es de longitud 1.

5. Pruébese que sobre un espacio unitario complejo se puede extraer la raíz de cualquier grado natural m de toda aplicación normal, es decir, que para toda aplicación normal $\mathcal{A}$ existe una aplicación $\mathcal{B}$, también normal, tal que $\mathcal{B}^m = \mathcal{A}$. ¿Cuál es el número máximo de estas aplicaciones $\mathcal{B}$?

6. Demuéstrese que siendo $\mathfrak{A}$ un subespacio invariante de una aplicación normal $\mathcal{A}$, el complemento ortogonal $\mathfrak{A}^\perp$ también será invariante respecto de $\mathcal{A}$. Si el espacio principal es complejo, la propiedad indicada caracteriza las aplicaciones normales.

§ 19. Aplicaciones unitarias y simétricas

19.1. Aplicaciones unitarias. Una aplicación isomorfa de un espacio unitario sobre sí mismo se llama aplicación *unitaria* de este espacio. Con más detalle: una aplicación lineal regular $\mathcal{U}$ de un espacio unitario $\mathfrak{L}$ se llama unitaria, si no altera el valor del producto escalar, es decir, si para todos los a y b de $\mathfrak{L}$ tiene lugar la relación

$$(a, b) = (a\mathcal{U}, b\mathcal{U}). \tag{1}$$

Las rotaciones del espacio euclideo corriente de tres dimensiones alrededor del origen de coordenadas O representan el ejemplo más sencillo de aplicaciones unitarias. Las reflexiones especulares de este espacio respecto a un plano cualquiera que pasa por O representan otro ejemplo de aplicaciones unitarias del espacio corriente. Se puede probar fácilmente que con las combinaciones de estos dos tipos de aplicaciones se agotan todas las aplicaciones unitarias del espacio corriente. Por esto las aplicaciones unitarias de espacios se pueden considerar como las aplicaciones análogas a las rotaciones y a las reflexiones especulares del espacio euclídeo corriente.

De la igualdad (1), que caracteriza las aplicaciones unitarias, se deduce que

$$(x,\ y) = (x\mathcal{U},\ y\mathcal{U}) = (x,\ y\mathcal{U}\mathcal{U}^*),$$

de donde resulta que

$$\mathcal{U}\mathcal{U}^* = \mathcal{E},\ \mathcal{U}^* = \mathcal{U}^{-1} \text{ y } \mathcal{U}^*\mathcal{U} = \mathcal{E}. \tag{2}$$

Recíprocamente, de las relaciones (2) se deduce que $\mathcal{U}$ es invertible y que

$$(x,\ y) = (x,\ y\mathcal{U}\mathcal{U}^*) = (x\mathcal{U},\ y\mathcal{U}).$$

Por consiguiente, *una aplicación lineal $\mathcal{U}$ es unitaria cuando, y solo cuando, la aplicación conjugada $\mathcal{U}^*$ coincide con la inversa $\mathcal{U}^{-1}$.*

En particular, las relaciones (2) muestran que *las aplicaciones unitarias son aplicaciones normales en el sentido del punto anterior.*

Tomemos en un espacio $\mathfrak{L}$ un sistema ortonormal de coordenadas y sea $\mathcal{U}$ una aplicación unitaria de este espacio. Si la matriz de la aplicación $\mathcal{U}$ es igual a U, la matriz de la aplicación conjugada es igual a $\overline{U}'$, de acuerdo con el p. 18.2. De la relación (2) se desprende por lo tanto que

$$U\overline{U}' = E. \tag{3}$$

Recíprocamente, si la matriz U de una aplicación lineal $\mathcal{U}$ satisface en un sistema ortonormal de coordenadas la relación (3), la propia aplicación $\mathcal{U}$ satisface las relaciones (2) y es, por consiguiente, unitaria. Las matrices U que satisfacen la relación (3) se llaman *unitarias* (p. 1.3); llegamos, por lo tanto, al resultado siguiente: *toda aplicación unitaria tiene en un sistema ortonormal de coordenadas una matriz unitaria; recíprocamente, si la matriz de una aplicación lineal es unitaria en un sistema ortonormal de coordenadas, la propia aplicación es unitaria.*

Siendo real el campo principal K las matrices de las aplicaciones también son reales y la relación (3) se convierte en

$$UU' = E. \tag{4}$$

Las matrices que satisfacen esta relación han sido llamados en el p. 1.3. ortogonales. Es decir, toda aplicación real unitaria tiene en una base ortonormal la matriz ortogonal. Recíprocamente, si en una base ortonormal la matriz de una aplicación lineal de un espacio real unitario es ortogonal, la aplicación es unitaria.

Las aplicaciones unitarias no alteran, por definición, los valores de los productos escalares. De aquí resulta que las aplicaciones unitarias no alteran las longitudes de los vectores.

La última propiedad es característica para las aplicaciones unitarias: *si una aplicación lineal $\mathcal{U}$ de un espacio unitario $\mathfrak{L}$ no altera las longitudes de los vectores, la aplicación $\mathcal{U}$ es unitaria.*

En efecto, sean a y b unos vectores arbitrarios de $\mathfrak{L}$. Tomemos

$$a\mathcal{U} = a' \text{ y } b\mathcal{U} = b'.$$

Puesto que la aplicación $\mathcal{U}$ es lineal, tenemos

$$(a + \alpha b)\,\mathcal{U} = a' + \alpha b'$$

para cualquier valor α del campo de coeficientes. Por hipótesis, la aplicación $\mathcal{U}$ no altera las longitudes y por lo tanto

$$(a + \alpha b,\ a + \alpha b) = (a' + \alpha b',\ a' + \alpha b').$$

Realizando aquí la multiplicación y reduciendo los términos semejantes, obtenemos

$$\alpha\,(b,\ a) + \bar{\alpha}\,(a,\ b) = \alpha\,(b',\ a') + \bar{\alpha}\,(a',\ b'). \tag{5}$$

Para $\alpha = 1$ esta igualdad se convierte en

$$(b,\ a) + (a,\ b) = (b',\ a') + (a',\ b'). \tag{6}$$

Si el campo principal es real, tenemos $(b,\ a) = (a,\ b)$ y de (6) resulta $(a,\ b) = (a',\ b')$. En cambio, si K no es real, tomando en (5) $\alpha = i$ y simplificando en i, llegamos a la relación

$$(b,\ a) - (a,\ b) = (b',\ a') - (a',\ b')$$

que con (6) da de nuevo $(a,\ b) = (a',\ b')$. Luego, tenemos en ambos casos

$$(a,\ b) = (a',\ b') = (a\mathcal{U},\ b\mathcal{U}),$$

es decir, la aplicación $\mathcal{U}$ es unitaria.

Examinemos el problema acerca de la transformación de coordenadas en los espacios unitarios. Sea $e_1,\ e_2,\ \ldots,\ e_n$ una base ortonormal de un espacio unitario $\mathfrak{L}$ y sea $\mathcal{U}$ una aplicación unitaria cualquiera de este espacio. Puesto que una aplicación unitaria no altera las longitudes de los vectores y transforma los vectores ortogonales en ortogonales, el sistema $e_1\mathcal{U},\ e_2\mathcal{U},\ \ldots,\ e_n\mathcal{U}$ será de nuevo una base ortonormal de $\mathfrak{L}$. Recíprocamente, supongamos que una aplicación lineal $\mathcal{U}$ transforma una base ortonormal $e_1,\ e_2,\ \ldots,\ e_n$ en una

base también ortonormal $e_1\mathcal{U}, e_2\mathcal{U}, \ldots, e_n\mathcal{U}$. Tomemos en $\mathfrak{L}$ unos vectores arbitrarios

$$a=\alpha_1e_1+\alpha_2e_2+\ldots+\alpha_ne_n \text{ y } b=\beta_1e_1+\beta_2e_2+\ldots+\beta_ne_n.$$

Tenemos

$$(a,\ b)=(\sum\alpha_ie_i,\ \sum\beta_ie_i)=\alpha_1\overline{\beta}_1+\alpha_2\overline{\beta}_2+\ldots+\alpha_n\overline{\beta}_n,$$
$$(a\mathcal{U},\ b\mathcal{U})=(\sum\alpha_ie_i\mathcal{U},\ \sum\beta_ie_i\mathcal{U})=\alpha_1\overline{\beta}_1+\alpha_2\overline{\beta}_2+\ldots+\alpha_n\overline{\beta}_n,$$

es decir, $(a,\ b)=(a\mathcal{U},\ b\mathcal{U})$; luego, la aplicación $\mathcal{U}$ es unitaria. Por consiguiente, *para que una aplicación lineal $\mathcal{U}$ sea unitaria es necesario y suficiente que $\mathcal{U}$ transforme una base ortonormal en una base ortonormal.*

De aquí también se desprende directamente la proposición siguiente: *la matriz del cambio de una base ortonormal por otra es unitaria y, viceversa, si una de las bases es ortonormal y la matriz del cambio es unitaria, la otra base es también ortonormal.*

Para la demostración basta observar que la matriz del cambio de un sistema de coordenadas por otro coincide con la matriz de la aplicación lineal que transforma el primer sistema en el segundo (p. 8.3).

Notemos, finalmente, que de la definición de las aplicaciones unitarias se deducen fácilmente las siguientes propiedades de las mismas: 1) *la aplicación idéntica es unitaria*, 2) *el producto de aplicaciones unitarias es unitario* y 3) *la inversa de una aplicación unitaria es una aplicación unitaria.*

19.2. Equivalencia unitaria. Ligado de un modo natural al concepto de aplicación unitaria aparece uno de los problemas principales de la teoría de espacios unitarios. Se trata de la clasificación de las aplicaciones lineales de estos espacios. Sean $\mathfrak{L}$ y $\mathfrak{L}_1$ unos espacios unitarios sobre un mismo campo principal K. Consideremos dos aplicaciones lineales $\mathcal{A}$ y $\mathcal{A}_1$ de estos espacios con la particularidad de que $\mathcal{A}$ actúa sobre $\mathfrak{L}$, mientras que $\mathcal{A}_1$ actúa sobre $\mathfrak{L}_1$. Las aplicaciones $\mathcal{A}$ y $\mathcal{A}_1$ se llaman *semejantes o isomorfas* si existe una aplicación isomorfa de $\mathfrak{L}$ sobre $\mathfrak{L}_1$ que transforme la aplicación $\mathcal{A}$ en la aplicación $\mathcal{A}_1$. Puesto que todos los espacios unitarios son, salvo un isomorfismo, conocidos y se determinan por su dimensión n, podemos aceptar que $\mathfrak{L}_1$ coincide con $\mathfrak{L}$ y, que, por consiguiente, las aplicaciones $\mathcal{A}$ y $\mathcal{A}_1$ actúan sobre el mismo espacio $\mathfrak{L}$. En este caso nuestra definición significa que $\mathcal{A}$ y $\mathcal{A}_1$ son isomorfas cuando, y sólo cuando, existe una aplicación isomorfa $\mathcal{U}$ del espacio $\mathfrak{L}$ sobre sí mismo que transforma $\mathcal{A}$ en $\mathcal{A}_1$. Como hemos visto en el p. 10.2 esto equivale a la condición

$$\mathcal{A}_1=\mathcal{U}^{-1}\mathcal{A}\mathcal{U}. \tag{7}$$

Si tomamos en $\mathfrak{L}$ un sistema ortonormal de coordenadas, la relación (7)

puede ser representada en la forma matricial

$$A_1 = U^{-1}AU, \tag{8}$$

donde U es una matriz unitaria y A y A_1 son las matrices de las aplicaciones lineales dadas. Unas matrices A y A_1 que satisfacen la relación (8) se llaman *unitariamente equivalentes*. Por consiguiente, *las aplicaciones lineales de un espacio unitario son isomorfas cuando, y sólo cuando, sus matrices, calculadas en una base ortonormal, son unitariamente equivalentes.*

Considerando la matriz U de la relación (8) como una matriz de cambio, llegamos a la siguiente proposición: *unas aplicaciones lineales $\mathcal{A}$ y $\mathcal{B}$ de un espacio unitario $\mathfrak{L}$ son isomorfas cuando, y sólo cuando, en $\mathfrak{L}$ existen dos bases ortonormales tales que la matriz de la aplicación $\mathcal{A}$ calculada en una de ellas coincide con la matriz de la aplicación $\mathcal{B}$ calculada en la otra.*

Esta proposición es totalmente análoga a la correspondiente afirmación para espacios lineales arbitrarios que ha sido considerada en el p. 10.2, donde se ha dado también una demostración detallada.

A título de ejemplo, podemos tomar las aplicaciones normales. Está claro que para que unas aplicaciones lineales de un espacio unitario sean *unitariamente* isomorfas es necesario que sean isomorfas *linealmente*, es decir, que sean isomorfas como aplicaciones lineales de un espacio lineal. Por ello, para que unas aplicaciones lineales sean unitariamente isomorfas es necesario que sus polinomios característicos coincidan. Si las aplicaciones dadas son normales y las raíces de sus polinomios característicos son conocidas, es posible, según el punto anterior, escribir las matrices de estas aplicaciones en unas bases ortonormales del espacio convenientemente escogidas. Puesto que estas matrices resultarán idénticas, las aplicaciones serán unitariamente isomorfas. Hemos demostrado, pues, el siguiente teorema:

TEOREMA 1. *Para que unas aplicaciones normales de espacios unitarios, tanto reales como complejos, sean unitariamente isomorfas es necesario y suficiente que los polinomios característicos de estas aplicaciones coincidan.*

En una forma puramente matricial el teorema 1 puede ser enunciado del siguiente modo:

TEOREMA 1a. *Para la equivalencia unitaria de unas matrices A y B que conmutan con sus matrices anticonjugadas complejas $\bar{A}$ y $\bar{B}$ es necesario y suficiente, tanto en el caso del cuerpo de los números complejos como en el caso del cuerpo de los números reales, que los polinomios característicos de las matrices A y B coincidan.*

En particular, para toda matriz compleja o real A, tal que $A\bar{A}' = A'\bar{A}$, existe una matriz unitaria U compleja o, respectivamente, real tal que la matriz $UAU^{-1} = UA\overline{U}'$ es diagonal en el caso complejo y tiene la forma (16) del p. 18.3 en el caso real.

Por consiguiente, el isomorfismo unitario de las aplicaciones normales resulta ser equivalente al isomorfismo corriente de las mismas.

19.3. Forma normal de la matriz de una aplicación unitaria. Como ya hemos señalado, las aplicaciones unitarias son un caso particular de las normales. Por esto los teoremas 1 y 1a, además de ofrecer las condiciones necesarias y suficientes para que un isomorfismo sea unitario, ofrecen también la forma normal para las matrices de las aplicaciones unitarias. El teorema que sigue indica el rasgo característico que distingue las aplicaciones unitarias de las demás aplicaciones normales:

TEOREMA 2. *El módulo de todas las raíces del polinomio característico de una aplicación unitaria es igual a la unidad.*

Consideremos primero el caso complejo. En este caso a toda raíz α del polinomio característico de una aplicación unitaria $\mathcal{U}$ le corresponde un vector propio no nulo a. De las relaciones

$$a\mathcal{U} = \alpha a \text{ y } (a\mathcal{U}, a\mathcal{U}) = (a, a)$$

resulta

$$(\alpha a, \alpha a) = \alpha\bar{\alpha}(a, a) = (a, a),$$

es decir, $\alpha\bar{\alpha} = |\alpha|^2 = 1$.

En el caso real, a toda raíz compleja $\alpha = \rho + i\sigma$ le corresponde un par a y b de vectores ortogonales no nulos tales que

$$a\mathcal{U} = \rho a - \sigma b \text{ y } b\mathcal{U} = \sigma a + \rho b.$$

De aquí tenemos

$$(a, a) + (b, b) = (a\mathcal{U}, a\mathcal{U}) + (b\mathcal{U}, b\mathcal{U}) = (\rho^2 + \sigma^2)((a, a) + (b, b)),$$

es decir,

$$\rho^2 + \sigma^2 = |\alpha|^2 = 1.$$

Observando que entre los números reales sólo los números 1 y -1 son de módulo igual a la unidad, podemos enunciar el teorema 1a en el caso de aplicaciones unitarias en la forma siguiente:

TEOREMA 3. *Para toda matriz unitaria A existe una matriz unitaria compleja U tal que UAU^{-1} será una matriz diagonal con elementos diagonales de módulo igual a la unidad. Para toda matriz unitaria real A existe una matriz unitaria real U tal que*

$$UAU^{-1} = E_s + (-E_t) + \begin{bmatrix} \cos\varphi_1 & \operatorname{sen}\varphi_1 \\ -\operatorname{sen}\varphi_1 & \cos\varphi_1 \end{bmatrix} + \cdots$$

$$\cdots + \begin{bmatrix} \cos\varphi_m & \operatorname{sen}\varphi_m \\ -\operatorname{sen}\varphi_m & \cos\varphi_m \end{bmatrix}, \qquad (9)$$

donde E_s y E_t son matrices unidades de orden s y t, respectivamente, y $\operatorname{sen}\varphi_j \neq 0$ $(j = 1, \ldots, m)$, *con la particularidad de que algunos de los números s, t y m pueden ser iguales a cero, es decir, en la fórmula* (9) *pueden no figurar los términos correspondientes.*

Consideremos, a título de ejemplo, el espacio euclídeo corriente de tres dimensiones $\mathfrak{R}$. Para toda aplicación ortogonal $\mathcal{U}$ del espacio $\mathfrak{R}$ se puede encontrar, de acuerdo con el teorema 3, un sistema e_1, e_2 y e_3 de vectores perpendiculares unitarios tal que la matriz de la aplicación $\mathcal{U}$ tomará una de las seis formas siguientes:

$$\alpha)\begin{bmatrix}1&&\\&1&\\&&1\end{bmatrix},\quad \beta)\begin{bmatrix}-1&&\\&1&\\&&1\end{bmatrix},\quad \gamma)\begin{bmatrix}1&&\\&-1&\\&&-1\end{bmatrix},$$

$$\delta)\begin{bmatrix}-1&&\\&-1&\\&&-1\end{bmatrix},\quad \varepsilon)\begin{bmatrix}1&&\\&\cos\varphi&\operatorname{sen}\varphi\\&-\operatorname{sen}\varphi&\cos\beta\end{bmatrix}\quad \text{y}\quad \varkappa)\begin{bmatrix}-1&&\\&\cos\varphi&\operatorname{sen}\varphi\\&-\operatorname{sen}\varphi&\cos\varphi\end{bmatrix}.$$

Es evidente que la aplicación $\mathcal{U}$ es la aplicación idéntica en el caso α), la reflexión especular respecto al plano e_2Oe_3 en el caso β), la reflexión especular respecto a la recta Oe_1 en el caso γ), la reflexión especular respecto al origen O en el caso δ), la rotación de ángulo φ alrededor del eje e_1 en el caso ε) y la rotación de ángulo φ alrededor del eje e_1 seguida de la reflexión especular respecto al plano e_2Oe_3 en el caso $\varkappa$). Los cuatro primeros casos se pueden considerar como casos particulares de los dos últimos con $\varphi = 0$ y $\varphi = \pi$.

19.4. Aplicaciones simétricas. Una aplicación lineal $\mathcal{A}$ de un espacio unitario $\mathfrak{L}$ se llama *hermitiana o simétrica*, si $\mathcal{A}$ coincide con su aplicación conjugada $\mathcal{A}^*$. Es decir, si la aplicación $\mathcal{A}$ es simétrica, se tiene

$$(x\mathcal{A}, y) = (x, y\mathcal{A}). \tag{10}$$

Recíprocamente, si una aplicación lineal $\mathcal{A}$ satisface la condición (10) cualesquiera que sean x e y de $\mathfrak{L}$, la aplicación $\mathcal{A}$ es simétrica.

Es evidente que de la condición $\mathcal{A}^* = \mathcal{A}$ se deduce la igualdad $\mathcal{A}\mathcal{A}^* = \mathcal{A}^*\mathcal{A}$, es decir, *las aplicaciones simétricas son aplicaciones normales.*

Tomemos en $\mathfrak{L}$ una base ortonormal y sea A la matriz de una aplicación simétrica $\mathcal{A}$. La matriz de la aplicación conjugada $\mathcal{A}^*$ es igual en esta base a la matriz anticonjugada $\overline{A'}$. Tenemos, por hipótesis, $\mathcal{A}^* = \mathcal{A}$, de donde

$$\overline{A'} = A. \tag{11}$$

Recíprocamente, de (11) se deduce que $\mathcal{A}^* = \mathcal{A}$, es decir, que $\mathcal{A}$ es simétrica. Las matrices que satisfacen la relación (11) han sido

llamadas en el p. 1.3 hermitianas. Por consiguiente, *en una base ortonormal a las aplicaciones simétricas les corresponden matrices hermitianas y, viceversa, a las matrices hermitianas les corresponden las aplicaciones simétricas.*

El ejemplo más sencillo de aplicación simétrica es la aplicación de tipo $\alpha\mathscr{E}$, donde α es real. El ejemplo general se obtiene con las aplicaciones cuyas matrices tienen, en una base ortonormal, la forma diagonal real.

La suma de aplicaciones simétricas y el producto de una aplicación simétrica por un número real son de nuevo aplicaciones simétricas.

En efecto, si las aplicaciones $\mathscr{A}$ y $\mathscr{B}$ son simétricas y α es un número real, se tiene

$$(\mathscr{A}+\mathscr{B})^* = \mathscr{A}^* + \mathscr{B}^* = \mathscr{A} + \mathscr{B},$$

$$(\alpha\mathscr{A})^* = \bar{\alpha}\mathscr{A}^* = \alpha\mathscr{A}.$$

El producto de dos aplicaciones simétricas es una aplicación simétrica cuando, y sólo cuando, estas aplicaciones son conmutables.

Efectivamente, de $\mathscr{AB} = \mathscr{BA}$. $\mathscr{A} = \mathscr{A}^*$ y $\mathscr{B} = \mathscr{B}^*$ se deduce que

$$(\mathscr{AB})^* = \mathscr{B}^*\mathscr{A}^* = \mathscr{BA} = \mathscr{AB}.$$

Recíprocamente, si $(\mathscr{AB})^* = \mathscr{AB}$, $\mathscr{A} = \mathscr{A}^*$ y $\mathscr{B} = \mathscr{B}^*$, se tiene

$$\mathscr{AB} = (\mathscr{AB})^* = \mathscr{B}^*\mathscr{A}^* = \mathscr{BA}.$$

De aquí se deduce, en particular, que las potencias de una aplicación simétrica y, en general, los polinomios de coeficientes reales en una aplicación simétrica son de nuevo aplicaciones simétricas.

Los valores propios de las aplicaciones simétricas son reales.

Efectivamente, sea $\mathscr{A}$ una aplicación simétrica, sea α un valor propio de la misma y sea a un vector propio no nulo correspondiente. Tenemos

$$(a, a\mathscr{A}) = (a, \alpha a) = \bar{\alpha}(a, a),$$
$$(a\mathscr{A}, a) = (\alpha a, a) = \alpha(a, a).$$

Pero

$$(a, a\mathscr{A}) = (a\mathscr{A}, a).$$

Comparando estos resultados, vemos que $\bar{\alpha} = \alpha$, es decir, que α es real.

Todas las raíces del polinomio característico de una matriz hermitiana son reales.

En efecto, toda matriz hermitiana A puede ser considerada como la matriz de una aplicación simétrica $\mathscr{A}$ de un espacio unitario. Las raíces del polinomio característico de la matriz A son los valores propios de la aplicación $\mathscr{A}$ y, por consiguiente, son reales.

Hemos señalado anteriormente que las aplicaciones simétricas son normales. Por lo tanto, para que unas aplicaciones simétricas sean unitariamente isomorfas es necesario y suficiente, en virtud

del teorema 1, que los polinomios característicos de estas aplicaciones coincidan.

Para toda aplicación normal de un espacio unitario complejo existe, según el teorema 6 del p. 18.3, una base ortonormal en la que la matriz de la aplicación es de forma diagonal. Puesto que todos los valores propios de las aplicaciones simétricas son reales, la matriz diagonal indicada será real en el caso de aplicaciones simétricas; obtenemos así el teorema siguiente:

TEOREMA 4. *Toda aplicación simétrica de un espacio complejo unitario tiene, en una base ortonormal adecuada, una matriz diagonal real.*

La recíproca es también válida, ya que si la matriz de una aplicación lineal $\mathcal{A}$ en una base ortonormal es de forma diagonal real A, se tiene $\overline{A'} = A$ y, por consiguiente, $\mathcal{A}^* = \mathcal{A}$.

En términos matriciales el teorema 4 puede ser enunciado de la forma siguiente.

TEOREMA 4a. *Para toda matriz hermitiana A existe una matriz unitaria compleja U tal que la matriz UAU^{-1} es de forma diagonal real.*

Consideremos ahora el caso en que $\mathcal{A}$ es una aplicación simétrica de un espacio unitario real. Según el teorema 8 del p. 18.3, la matriz de la aplicación $\mathcal{A}$ se descompone, en una base ortonormal adecuada, en células de orden 1 ó 2. Además, las células de orden 2 aparecen sólo cuando el polinomio característico de la aplicación tiene raíces no reales. Pero los polinomios característicos de las aplicaciones simétricas no tienen raíces que no sean reales. Por consiguiente, también en el caso real la matriz de una aplicación simétrica se reduce a la forma diagonal en una base ortonormal adecuada. La recíproca, obviamente, es también válida, de modo que tiene lugar el teorema siguiente:

TEOREMA 5. *Para toda aplicación simétrica de un espacio unitario real existe una base ortonormal en la que la matriz de la aplicación adquiere la forma diagonal.*

En una base ortonormal las matrices de las aplicaciones simétricas satisfacen las relaciones $\overline{A'} = A$. En el caso real está relación se convierte en la igualdad $A' = A$. Las matrices que satisfacen esta igualdad se llaman simétricas (p. 1.3). Por consiguiente, en una base ortonormal las matrices de las aplicaciones simétricas reales son simétricas y, viceversa, las aplicaciones son simétricas si sus matrices son simétricas reales. Esta observación permite enunciar el teorema 5 del modo siguiente.

TEOREMA 5a. *Para toda matriz simétrica real A existe una matriz unitaria real U tal que la matriz UAU^{-1} es de forma diagonal.*

19.5. Aplicaciones antisimétricas. Sea $\mathfrak{L}$ un espacio unitario. Una aplicación lineal $\mathcal{A}$ se llama *antisimétrica*, si está ligada a su

aplicación conjugada por la relación

$$\mathcal{A}^* = -\mathcal{A}. \tag{12}$$

Siendo a y b vectores arbitrarios de $\mathfrak{L}$, de (12) se deduce que

$$(a, b\mathcal{A}) = (a\mathcal{A}^*, b) = -(a\mathcal{A}, b).$$

Recíprocamente, si para cualesquiera a y b se tiene

$$(a, b\mathcal{A}) = -(a\mathcal{A}, b),$$

resulta que $\mathcal{A} = -\mathcal{A}^*$ y $\mathcal{A}$ es una aplicación antisimétrica.

En el caso en el que el campo principal es el cuerpo de los números complejos, las aplicaciones antisimétricas tienen una expresión muy sencilla mediante las simétricas. En efecto, sea $\mathcal{A}$ una aplicación simétrica de un espacio $\mathfrak{L}$. Entonces

$$(i\mathcal{A})^* = \bar{i}\mathcal{A}^* = -i\mathcal{A},$$

es decir, la aplicación $i\mathcal{A}$ es antisimétrica. Recíprocamente, si la aplicación $\mathcal{A}$ es antisimétrica, se tiene

$$(i\mathcal{A})^* = \bar{i}\mathcal{A}^* = i\mathcal{A},$$

y, por consiguiente, la aplicación $i\mathcal{A}$ es simétrica. En los espacios reales esta relación desaparece.

Tomemos una base ortonormal en un espacio unitario $\mathfrak{L}$ e indiquemos por A la matriz de una aplicación antisimétrica $\mathcal{A}$. Puesto que la matriz de la aplicación conjugada $\mathcal{A}^*$ es igual a $\overline{A'}$, la condición (12) da

$$\overline{A'} = -A. \tag{13}$$

Viceversa, de (13) se deduce, obviamente, que $\mathcal{A}$ es una aplicación antisimétrica. Las matrices que verifican la relación (13) se llaman matrices *hermitianas antisimétricas*. Por consiguiente, en una base ortonormal a las aplicaciones antisimétricas les corresponden las matrices hermitianas antisimétricas y, viceversa, a las matrices hermitianas antisimétricas les corresponden aplicaciones antisimétricas.

De la relación (12) se deduce directamente que la suma de aplicaciones antisimétricas y el producto de una aplicación antisimétrica por un número real son de nuevo aplicaciones antisimétricas.

Todo valor propio de una aplicación antisimétrica o bien es igual a cero o bien es un número imaginario puro.

Si α es un valor propio de una aplicación antisimétrica $\mathcal{A}$ y a es un vector propio no nulo correspondiente, se tiene

$$(a, a\mathcal{A}) = (a, \alpha a) = \bar{\alpha}(a, a),$$
$$(a\mathcal{A}, a) = (\alpha a, a) = \alpha(a, a).$$

Pero $(a, a\mathcal{A}) = -(a\mathcal{A}, a)$ y, por consiguiente, $-\alpha = \bar{\alpha}$ que es lo que se quería demostrar.

En particular, de aquí resulta que *toda raíz del polinomio característico de una matriz hermitiana antisimétrica o bien es igual a cero o bien es un número imaginario puro.*

Puesto que de la relación $\mathcal{A}^* = -\mathcal{A}$ se deduce inmediatamente la igualdad $\mathcal{A}^*\mathcal{A} = \mathcal{A}\mathcal{A}^*$, *las aplicaciones antisimétricas resultan ser un caso particular de las aplicaciones normales* y, por ello, para hallar la forma más sencilla de las matrices de las aplicaciones antisimétricas es suficiente recurrir al teorema 8 del p. 18.3. Así se obtiene el teorema siguiente:

TEOREMA 6 *En una base ortonormal adecuada de un espacio unitario real la matriz A de una aplicación antisimétrica toma la forma*

$$A = O_k + \begin{bmatrix} 0 & \sigma_1 \\ -\sigma_1 & 0 \end{bmatrix} + \ldots + \begin{bmatrix} 0 & \sigma_m \\ -\sigma_m & 0 \end{bmatrix}, \qquad (14)$$

donde O_k es la matriz nula de orden k.

Efectivamente, en una base ortonormal adecuada la matriz de la aplicación dada se descompone, según el teorema mencionado, en células de órdenes 1 y 2. De la relación (13) se ve que también las células aisladas deben satisfacer esta misma igualdad. Las células de orden 1 son números reales ρ_j y la relación (13) da para ellas $\rho_j = -\overline{\rho}_j = -\rho_j$, es decir, $\rho_j = 0$. En cambio, si la célula es de orden 2, de (13) resulta que debe ser de la forma

$$A_j = \begin{bmatrix} 0 & \sigma_j \\ -\sigma_j & 0 \end{bmatrix}$$

que es lo que se quería demostrar.

En términos matriciales el teorema 6 se puede enunciar de modo siguiente:

TEOREMA 6a. *Para toda matriz antisimétrica real A existe una matriz unitaria real U tal que la matriz UAU^{-1} es de la forma* (14).

Para las matrices reales los conceptos de matriz unitaria y de matriz ortogonal son equivalentes y por esto en los teoremas 5a y 6a las palabras *unitaria real* pueden ser sustituidas por las palabras *unitaria ortogonal.*

19.6. Aplicaciones simétricas no negativas. Una aplicación simétrica $\mathcal{A}$ de un espacio unitario $\mathfrak{L}$ se llama *no negativa*, si para todo x de $\mathfrak{L}$ se tiene

$$(x\mathcal{A}, x) \geqslant 0. \qquad (15)$$

Aquí el signo de desigualdad tiene sentido, ya que en el caso de aplicaciones simétricas el producto escalar $(x\mathcal{A}, x)$ es siempre real. Si el signo de igualdad tiene lugar en (15) sólo para el vector nulo, se dice que $\mathcal{A}$ es una aplicación *positiva* o *definida positiva.*

Una combinación lineal de aplicaciones no negativas con coeficientes reales no negativos es una aplicación no negativa.

Esto se ve directamente de la fórmula

$$(x(\alpha\mathcal{A}+\beta\mathcal{B}),\ x)=\alpha(x\mathcal{A},\ x)+\beta(x\mathcal{B},\ x).$$

El producto de cualquier aplicación lineal por su conjugada es una aplicación simétrica no negativa.

Efectivamente,

$$(x\mathcal{A}\mathcal{A}^*,\ x)=(x\mathcal{A},\ x\mathcal{A})\geqslant 0,$$
$$(x\mathcal{A}^*\mathcal{A},\ x)=(x\mathcal{A}^*,\ x\mathcal{A}^*)\geqslant 0.$$

El cuadrado de cualquier aplicación simétrica es una aplicación no negativa.

Resulta de la anterior, ya que toda aplicación simétrica es conjugada de sí misma.

Todos los valores propios de una aplicación no negativa son reales y no negativos.

Sea $\mathcal{A}$ una aplicación no negativa, sea α un valor propio de la misma y sea a un vector propio no nulo correspondiente. Entonces

$$(a\mathcal{A},\ a)=\alpha(a,\ a)\geqslant 0.$$

Por consiguiente, se tiene $\alpha\geqslant 0$.

Si los valores propios de una aplicación simétrica de un espacio unitario Ω, complejo o real, son no negativos, la aplicación es no negativa.

En Ω existe, en virtud del p. 19.4, una base ortonormal $e_1, \ldots, e_n$ formada por los vectores propios de la aplicación $\mathcal{A}$. Sean $\alpha_1, \ldots, \alpha_n$ los valores propios correspondientes y sea

$$x=\xi_1e_1+\xi_2e_2+\ldots+\xi_ne_n$$

un vector arbitrario de Ω. Entonces

$$(x\mathcal{A},\ x)=\alpha_1\xi_1\bar{\xi}_1+\alpha_2\xi_2\bar{\xi}_2+\ldots+\alpha_n\xi_n\bar{\xi}_n=$$
$$=\alpha_1|\xi_1|^2+\ldots+\alpha_n|\xi_n|^2\geqslant 0 \qquad (16)$$

que es lo que se quería demostrar.

El determinante de la aplicación $\mathcal{A}$ es igual a $\alpha_1\alpha_2\ldots\alpha_n$. Si él es diferente de cero, todos los números α_j son mayores que el cero y la suma (16) será igual a cero en este caso sólo para $x=o$. Por consiguiente, la aplicación $\mathcal{A}$ será en este caso definida positiva. En cambio, si $|\mathcal{A}|=0$, uno de los valores propios, digamos α_1, es igual a cero. Entonces

$$(e_1\mathcal{A},\ e_1)=0(e_1,\ e_1)=0$$

y la aplicación $\mathcal{A}$ no será definida positiva.

Por consiguiente, *una aplicación simétrica no negativa es definida positiva cuando, y sólo cuando, es regular.*

Consideremos ahora la operación de extracción de la raíz cuadrada de una aplicación lineal. Se dice que una aplicación lineal $\mathcal{X}$

es raíz cuadrada de una aplicación lineal $\mathcal{A}$, si

$$\mathcal{X}^2 = \mathcal{A}. \tag{17}$$

Según sea la aplicación $\mathcal{A}$, la ecuación (17) puede no tener soluciones, puede tener sólo un número finito de soluciones y puede tener un número infinito de soluciones. Sin embargo, en el caso de aplicaciones simétricas no negativas la situación es bien determinada.

TEOREMA 7. *Para toda aplicación simétrica no negativa $\mathcal{A}$ de un espacio unitario existe una aplicación simétrica no negativa $\mathcal{B}$, y sólo una, que cumple la relación*

$$\mathcal{B}^2 = \mathcal{A}.$$

Toda aplicación lineal que conmuta con $\mathcal{A}$ es conmutable con $\mathcal{B}$.

DEMOSTRACIÓN. Tomemos en $\mathfrak{L}$ una base ortonormal $e_1, \ldots, e_n$ formada por los vectores propios de la aplicación $\mathcal{A}$. Tal base existe de acuerdo con el p. 19.4. Indiquemos por $\alpha_1, \ldots, \alpha_n$ los valores propios correspondientes de la aplicación $\mathcal{A}$. Sea $\mathcal{B}$ la aplicación lineal que transforma e_j en $\sqrt{\alpha_j}e_j$ $(j = 1, \ldots, n)$, donde se toman los valores no negativos de los radicales. Puesto que $e_1, \ldots, e_n$ es una base ortonormal formada por los vectores propios de la aplicación $\mathcal{B}$ y puesto que los valores propios de ésta son iguales a $\sqrt{\alpha_1}, \ldots, \sqrt{\alpha_n}$, es decir, son no negativos, resulta que $\mathcal{B}$ es una aplicación simétrica no negativa. Pero

$$e_j\mathcal{B}^2 = \alpha_j e_j = e_j\mathcal{A} \qquad (j = 1, \ldots, n).$$

Por lo tanto, $\mathcal{B}^2 = \mathcal{A}$. Hemos demostrado que la raíz cuadrada requerida de $\mathcal{A}$ existe.

Demostremos la última afirmación del teorema. Sea $\mathcal{X}$ una aplicación lineal que conmuta con $\mathcal{A}$. Tomemos los vectores coordenados $e_1, \ldots, e_n$ en tal orden que los valores propios iguales, si es que existen, correspondan a vectores coordenados adyacentes. Entonces las matrices de las aplicaciones $\mathcal{A}$ y $\mathcal{B}$ serán, respectivamente, de la forma

$$A = \begin{bmatrix} \alpha_1 E_1 & & & \\ & \alpha_2 E_2 & & \\ & & \ddots & \\ & & & \alpha_s E_s \end{bmatrix} \text{ y}$$

$$B = \begin{bmatrix} \sqrt{\alpha_1} E_1 & & & \\ & \sqrt{\alpha_2} E_2 & & \\ & & \ddots & \\ & & & \sqrt{\alpha_s} E_s \end{bmatrix},$$

donde $\alpha_1, \ldots, \alpha_s$ son *distintos* valores propios de la aplicación $\mathcal{A}$ y $E_1, \ldots, E_s$ son matrices unidades. Representemos la matriz de la aplicación $\mathcal{X}$ en la forma celular correspondiente

$$X = \begin{vmatrix} X_{11} & X_{12} & \ldots & X_{1s} \\ X_{21} & X_{22} & \ldots & X_{2s} \\ \ldots & \ldots & \ldots & \ldots \\ X_{s1} & X_{s2} & \ldots & X_{ss} \end{vmatrix}.$$

De la condición $AX = XA$ tenemos

$$\alpha_j X_{jk} = X_{jk}\alpha_k \qquad (j, k = 1, 2, \ldots, s),$$

es decir,

$$(\alpha_j - \alpha_k) X_{jk} = 0.$$

Puesto que $\alpha_j \neq \alpha_k$ para $j \neq k$, tenemos $X_{jk} = 0$ $(j \neq k)$. Por consiguiente,

$$X = \begin{vmatrix} X_{11} & 0 & \ldots & 0 \\ 0 & X_{22} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & X_{ss} \end{vmatrix};$$

pero entonces

$$BX = \begin{bmatrix} \sqrt{\alpha_1} X_{11} & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \sqrt{\alpha_s} X_{ss} \end{bmatrix} = XB$$

que es lo que se quería demostrar.

Resta demostrar la unicidad. Sea $\mathcal{C}$ otra aplicación simétrica no negativa tal que $\mathcal{C}^2 = \mathcal{A}$. Correspondientemente a la descomposición celular de la matriz A señalada anteriormente, el espacio $\mathfrak{L}$ se descompondrá en la suma directa de los subespacios invariantes $\mathfrak{L}_j$ $(j = 1, \ldots, s)$. Puesto que $\mathcal{C}\mathcal{A} = \mathcal{A}\mathcal{C}$, los subespacios $\mathfrak{L}_j$ serán también invariantes, de acuerdo con lo demostrado, respecto a $\mathcal{C}$. Como $\mathcal{C}$ es simétrica, en cada uno de los subespacios $\mathfrak{L}_j$ existe una base ortonormal formada por los vectores propios de $\mathcal{C}$. Sean $\gamma_{j1}, \ldots, \gamma_{jp}$ los valores propios correspondientes. Indiquemos por $\mathcal{A}_j$, $\mathcal{B}_j$ y $\mathcal{C}_j$ las aplicaciones inducidas en el subespacio $\mathfrak{L}_j$ por las aplicaciones $\mathcal{A}$, $\mathcal{B}$ y $\mathcal{C}$, respectivamente, tenemos

$$\mathcal{A}_j = \alpha_j \mathcal{E}_j, \quad \mathcal{B}_j = \sqrt{\alpha_j}\,\mathcal{E}_j \quad \text{y} \quad \mathcal{C}_j^2 = \gamma_j \mathcal{E}_j,$$

de donde

$$\gamma_{j1}^2 = \ldots = \gamma_{jp}^2 = \alpha_j.$$

Puesto que todos los números $\gamma_{j_1}, \ldots, \gamma_{j_p}$ son no negativos, de aquí se deduce que

$$\gamma_{j_1} = \ldots = \gamma_{j_p} = \sqrt{\alpha_j}$$

y, por consiguiente, $\mathscr{C}_j = \sqrt{\alpha_j}\mathscr{E}_j = \mathscr{B}_j$, es decir, $\mathscr{C} = \mathscr{B}$.

Demostremos, como ejemplo de una aplicación directa del teorema 7, que *el producto de aplicaciones simétricas no negativas conmutables es una aplicación no negativa.*

En efecto, sean $\mathscr{A}_1$ y $\mathscr{A}_2$ unas aplicaciones simétricas no negativas que conmutan. Indiquemos por $\mathscr{B}_1$ y $\mathscr{B}_2$ sus raíces cuadradas que, según el teorema 7, se pueden escoger de manera que conmuten y que sean simétricas y no negativas. Tenemos entonces

$$(\mathscr{B}_1\mathscr{B}_2)^2 = \mathscr{B}_1\mathscr{B}_2\mathscr{B}_1\mathscr{B}_2 = \mathscr{B}_1^2\mathscr{B}_2^2 = \mathscr{A}_1\mathscr{A}_2,$$

es decir, $\mathscr{A}_1\mathscr{A}_2$ es igual al cuadrado de una aplicación simétrica $\mathscr{B}_1\mathscr{B}_2$. Luego, $\mathscr{A}_1\mathscr{A}_2$ es no negativa. Hemos demostrado la proposición.

De ella se desprende, en particular, que los polinomios con coeficientes reales no negativos en una aplicación simétrica no negativa son también aplicaciones simétricas no negativas.

Ejemplos y problemas

1. Sea e_1, e_2 y e_3 una base ortonormal de un espacio unitario $\mathfrak{L}$. Hállense las matrices de las aplicaciones unitarias que transforman los vectores e_1 y e_2 en los vectores $\frac{2}{3}e_1 + \frac{2}{3}e_2 - \frac{1}{3}e_3$ y $\frac{2}{3}e_1 - \frac{1}{3}e_2 + \frac{2}{3}e_3$.

2. Si $a_1, \ldots, a_m$ y $b_1, \ldots, b_m$ son dos sistemas ortonormales de vectores de un espacio unitario $\mathfrak{L}$ de n dimensiones ($m \leqslant n$), existe una aplicación unitaria de $\mathfrak{L}$ que transforma el primer sistema en el segundo.

3. Para que un sistema de vectores $a_1, \ldots, a_m$ de un espacio unitario pueda ser transformado por una aplicación unitaria en otro sistema $b_1, \ldots, b_m$ es necesario y suficiente que las matrices de Gram de estos sistemas coincidan (véase el problema 5, pág. 215).

4. Demuéstrese, empleando la forma normal de Jordan y el proceso de ortonormalización de Gram—Schmidt, que la matriz de toda aplicación lineal de un espacio unitario puede ser reducida a la forma triangular en un sistema de coordenadas ortonormal adecuado (teorema de Schur).

5. Demuéstrese que siendo $\mathscr{A}$ una aplicación arbitraria de un espacio unitario $\mathfrak{L}$ que conserva los valores de los productos escalares, la aplicación $\mathscr{A}$ es lineal y es, por consiguiente, una aplicación unitaria del espacio $\mathfrak{L}$.

6. En un espacio lineal existen, salvo un isomorfismo, sólo dos funciones lineales. En un espacio unitario las funciones lineales son, salvo un isomorfismo, de la forma $\alpha(x, e)$, donde e es un vector unitario fijo.

7. En una base ortonormal de un espacio euclídeo las matrices de las aplicaciones $\mathscr{A}$, $\mathscr{B}$ y $\mathscr{C}$ son iguales respectivamente a

$$\begin{bmatrix} 5 & 2 & 4 \\ 2 & 2 & 2 \\ 4 & 2 & 5 \end{bmatrix}, \quad \begin{bmatrix} 4 & 2 & 4 \\ 2 & 1 & 2 \\ 4 & 2 & 4 \end{bmatrix} \text{ y } \begin{bmatrix} 1 & 3 & 1 \\ -1 & 2 & 2 \\ 1 & 0 & 2 \end{bmatrix}.$$

Demuéstrese que $\mathcal{A}$ es definida positiva, que $\mathcal{B}$ es no negativa y que $\mathcal{C}$, aun no siendo simétrica, es tal que $(x\mathcal{C}, x) \geqslant 0$ cualquiera que sea x.

8. Hállese la raíz cuadrada simétrica no negativa de la aplicación $\mathcal{B}$ del problema anterior.

9. La matriz de una aplicación simétrica no negativa, calculada en una base ortonormal, se llama *hermitiana no negativa*. Pruébese que una matriz hermitiana es no negativa cuando, y sólo cuando, se alternan los signos de los coeficientes de su polinomio característico. Además, si uno de los coeficientes es igual a cero, también tienen que ser iguales a cero los coeficientes de los términos de grado menor.

10. Demuéstrese que de toda aplicación normal $\mathcal{A}$ se puede extraer la raíz de cualquier grado positivo n, es decir, demuéstrese que para toda aplicación normal $\mathcal{A}$ existe una aplicación normal $\mathcal{X}$ que satisface la relación $\mathcal{X}^n = \mathcal{A}$. ¿Cuál es el número máximo de tales aplicaciones $\mathcal{X}$?

11. En un espacio unitario $\mathfrak{L}$ se ha tomado una base ortonormal. Demuéstrese que en esta base la matriz de toda aplicación simétrica no negativa $\mathcal{A}$ de rango 1 puede ser representada en la forma

$$A = \overline{[x]}^t [x],$$

donde $[x]$ es la fila coordenada de un vector x convenientemente escogido.

12. Toda aplicación simétrica no negativa es una suma de aplicaciones simétricas no negativas de rango 1.

13. Si unas matrices hermitianas de elementos α_{ij} y β_{ij} son no negativas, la matriz de elementos $\gamma_{ij} = \alpha_{ij}\beta_{ij}$ es también no negativa ($i, j = 1, 2, \ldots, n$).

§ 20. Descomposición de aplicaciones generales

Las aplicaciones unitarias, simétricas y antisimétricas tienen una estructura geométrica muy clara. Por esto al estudiar las aplicaciones lineales generales de espacios euclídeos o unitarios resulta natural preguntarse si es posible expresar de algún modo simple estas aplicaciones en términos de las aplicaciones especiales mencionadas. Algunos de estos métodos, que son de importancia principal, se consideran precisamente en este parrágrafo.

20.1. Descomposición en partes simétrica y antisimétrica. Sea $\mathfrak{L}$ un espacio unitario complejo y sea $\mathcal{A}$ una aplicación lineal del mismo. Designemos

$$\mathcal{B} = \frac{1}{2}(\mathcal{A} + \mathcal{A}^*) \quad \text{y} \quad \mathcal{C} = \frac{1}{2i}(\mathcal{A} - \mathcal{A}^*). \tag{1}$$

Tenemos

$$\mathcal{B}^* = \frac{1}{2}(\mathcal{A}^* + \mathcal{A}) = \mathcal{B} \quad \text{y} \quad \mathcal{C}^* = -\frac{1}{2i}(\mathcal{A}^* - \mathcal{A}) = \mathcal{C}.$$

Por consiguiente, $\mathcal{B}$ y $\mathcal{C}$ son simétricas. De (1) resulta

$$\mathcal{A} = \mathcal{B} + i\mathcal{C}. \tag{2}$$

Es decir, *toda aplicación lineal $\mathcal{A}$ de un espacio unitario complejo puede ser representada en la forma* (2), *donde $\mathcal{B}$ y $\mathcal{C}$ son aplicaciones simétricas*. Esta representación es unívoca, ya que de (2) se deduce que

$$\mathcal{A}^* = \mathcal{B}^* - i\mathcal{C}^* = \mathcal{B} - i\mathcal{C},$$

de donde obtenemos para $\mathcal{B}$ y $\mathcal{C}$ de nuevo las expresiones (1).

Si el campo principal es real, la descomposición (2) no sirve. En este caso se procede del modo siguiente. Sea $\mathcal{A}$ una aplicación arbitraria. Pongamos

$$\mathcal{B} = \frac{1}{2}(\mathcal{A} + \mathcal{A}^*) \quad \text{y} \quad \mathcal{C} = \frac{1}{2}(\mathcal{A} - \mathcal{A}^*). \tag{3}$$

De aquí resulta

$$\mathcal{A} = \mathcal{B} + \mathcal{C}. \tag{4}$$

Puesto que

$$\mathcal{B}^* = \frac{1}{2}(\mathcal{A}^* + \mathcal{A}) = \mathcal{B} \quad \text{y} \quad \mathcal{C}^* = \frac{1}{2}(\mathcal{A}^* - \mathcal{A}) = -\mathcal{C},$$

en la descomposición (4) $\mathcal{B}$ es una aplicación simétrica y $\mathcal{C}$ es una aplicación antisimétrica. La descomposición (4) es unívoca, ya que de ella se deduce que $\mathcal{A}^* = \mathcal{B} - \mathcal{C}$ y de aquí obtenemos para $\mathcal{B}$ y $\mathcal{C}$ de nuevo las expresiones (3). Por consiguiente, *toda aplicación lineal puede ser representada como la suma de una aplicación simétrica y una aplicación antisimétrica. Esta representación es unívoca.*

Es evidente que la descomposición (4) sirve cualquiera que sea el campo principal. La aplicación $\mathcal{B}$ se llama parte *simétrica* y la aplicación $\mathcal{C}$ se llama parte *antisimétrica* de la aplicación $\mathcal{A}$.

Desde el punto de vista del cálculo de matrices, la descomposición (2) significa que toda matriz cuadrada A se puede representar en la forma $B + iC$, donde B y C son matrices hermitianas, mientras que la descomposición (4) significa que toda matriz cuadrada A puede ser representada en la forma $B + C$, donde B es una matriz simétrica y C es una matriz antisimétrica.

20.2. Descomposición polar. Desde el punto de vista geométrico es de mucho mayor interés la representación de una aplicación lineal como el producto de unas aplicaciones simétrica y unitaria. La posibilidad de tal representación se basa en el lema siguiente.

LEMA. *Si las aplicaciones lineales $\mathcal{A}$ y $\mathcal{B}$ de un espacio unitario $\mathfrak{L}$ alteran igualmente las longitudes de los vectores*[1], *es decir, si para cualquier a*

$$(a\mathcal{A},\ a\mathcal{A}) = (a\mathcal{B},\ a\mathcal{B}), \tag{5}$$

existe una aplicación unitaria $\mathcal{U}$ del espacio $\mathfrak{L}$ tal que $\mathcal{A}\mathcal{U} = \mathcal{B}$.

Consideremos el dominio de valores de la aplicación $\mathcal{A}$, es decir, el conjunto de vectores de tipo $x\mathcal{A}$, donde x recorre todo el espacio $\mathfrak{L}$. Indiquemos por $\mathfrak{A}$ este dominio. Análogamente, indiquemos por $\mathfrak{B}$ el dominio de valores de la aplicación $\mathcal{B}$. $\mathfrak{A}$ y $\mathfrak{B}$ son, según el p. 10.1, unos subespacios lineales. Queremos, ante todo, establecer una correspondencia isomorfa entre $\mathfrak{A}$ y $\mathfrak{B}$. Sea a un vector de $\mathfrak{A}$.

[1] Las aplicaciones $\mathcal{A}$ y $\mathcal{B}$ que poseen esta propiedad se llaman *métricamente iguales*.

Busquemos en $\mathfrak{L}$ un vector x tal que $x\mathcal{A}=a$ y tomemos $x\mathcal{B}=b$. Convengamos en decir que b es la imagen de a y en indicarla por $a\mathcal{V}^0$. Probemos que a determina unívocamente a b. En efecto, lo contrario puede acontecer sólo si el vector x no se determina unívocamente por la condición $x\mathcal{A}=a$. Sin embargo, si x_1 es otro vector tal que $x_1\mathcal{A}=a$, tenemos $(x-x_1)\mathcal{A}=o$. De aquí se deduce, en virtud de (5), que

$$((x-x_1)\mathcal{B},\ (x-x_1)\mathcal{B})=((x-x_1)\mathcal{A},\ (x-x_1)\mathcal{A})=0,$$

es decir, que $(x-x_1)\mathcal{B}=o$ ó $x_1\mathcal{B}=x\mathcal{B}$ que es lo que se quería demostrar. Hemos probado de esta forma que $\mathcal{V}^0$ es una aplicación unívoca de $\mathfrak{A}$ en $\mathfrak{B}$. Sin embargo, es fácil ver que $\mathcal{V}^0$ es una aplicación *biyectiva* de $\mathfrak{A}$ sobre $\mathfrak{R}$. Efectivamente, siendo b un vector de $\mathfrak{B}$, existe en $\mathfrak{L}$ un vector x tal que $x\mathcal{B}=b$. Tomando entonces $x\mathcal{A}=a$, tendremos $a\mathcal{V}^0=b$ que es lo que se quería demostrar.

De la misma definición de la correspondencia $\mathcal{V}^0$ se desprende que para todo vector x de $\mathfrak{L}$ es válida la igualdad

$$x\mathcal{A}\mathcal{V}^0=x\mathcal{B}. \tag{6}$$

Empleando esta igualdad se puede probar que $\mathcal{V}^0$ es una aplicación *isomorfa* de $\mathfrak{A}$ sobre $\mathfrak{B}$. Efectivamente, sean a_1 y a_2 unos vectores de $\mathfrak{A}$. Busquemos en $\mathfrak{L}$ unos vectores x_1 y x_2 tales que $x_1\mathcal{A}=a_1$ y $x_2\mathcal{A}=a_2$. Tenemos entonces

$$(\alpha a_1+\beta a_2)\mathcal{V}^0=(\alpha x_1\mathcal{A}+\beta x_2\mathcal{A})\mathcal{V}^0=(\alpha x_1+\beta x_2)\mathcal{A}\mathcal{V}^0=(\alpha x_1+\beta x_2)\mathcal{B}=$$
$$=\alpha(x_1\mathcal{B})+\beta(x_2\mathcal{B})=\alpha(x_1\mathcal{A}\mathcal{V}^0)+\beta(x_2\mathcal{A}\mathcal{V}^0)=\alpha a_1\mathcal{V}^0+\beta a_2\mathcal{V}^0, \tag{7}$$

es decir, la aplicación $\mathcal{V}^0$ conserva las operaciones de adición y de multiplicación por número. Además,

$$(a_1\mathcal{V}^0,\ a_1\mathcal{V}^0)=(x_1\mathcal{A}\mathcal{V}^0,\ x_1\mathcal{A}\mathcal{V}^0)=(x_1\mathcal{B},\ x_1\mathcal{B})=$$
$$=(x_1\mathcal{A},\ x_1\mathcal{A})=(a_1,\ a_1), \tag{8}$$

es decir, la aplicación $\mathcal{V}^0$ conserva las longitudes de los vectores. Por consiguiente, $\mathcal{V}^0$ es un isomorfismo.

Tenemos definida la aplicación $\mathcal{V}^0$ sólo para los vectores de $\mathfrak{A}$. Queremos ahora definirla también en todos los demás vectores del espacio $\mathfrak{L}$. Con este fin consideremos los subespacios ortogonales $\mathfrak{A}^\perp$ y $\mathfrak{B}^\perp$. Para $\mathfrak{L}$ son válidas, de acuerdo con el p. 17.5, las descomposiciones directas

$$\mathfrak{L}=\mathfrak{A}\dotplus\mathfrak{A}^\perp=\mathfrak{B}\dotplus\mathfrak{B}^\perp.$$

Los subespacios $\mathfrak{A}$ y $\mathfrak{B}$ son isomorfos y, por lo tanto, tienen la misma dimensión. Pero entonces los complementos ortogonales $\mathfrak{A}^\perp$ y $\mathfrak{B}^\perp$ también tienen la misma dimensión. Como los espacios unitarios de una misma dimensión son isomorfos, debe existir una aplicación biyectiva de $\mathfrak{A}^\perp$ sobre $\mathfrak{B}^\perp$ que conserva las operaciones

de adición y de multiplicación por número y que no altera las longitudes de los vectores. Indiquemos esta aplicación por $\mathscr{W}$. Luego, si a' y a'' son unos vectores de $\mathfrak{A}^{\perp}$, los vectores $a'\mathscr{W}$ y $a''\mathscr{W}$ pertenecen a $\mathfrak{B}^{\perp}$ y

$$(\alpha a' + \beta a'')\mathscr{W} = \alpha(a'\mathscr{W}) + \beta(a''\mathscr{W}), \tag{9}$$

$$(a'\mathscr{W}, a''\mathscr{W}) = (a', a''). \tag{10}$$

Definamos ahora una aplicación $\mathscr{U}$ del espacio $\mathfrak{L}$ del modo siguiente. Sea x un vector cualquiera de $\mathfrak{L}$. Puesto que $\mathfrak{L} = \mathfrak{A} + \mathfrak{A}^{\perp}$, el vector x se puede representar unívocamente en la forma

$$x = x' + x'' \quad (x' \in \mathfrak{A} \quad \text{y} \quad x'' \in \mathfrak{A}^{\perp}). \tag{11}$$

Tomemos, por definición,

$$x\mathscr{U} = x'\mathscr{V} + x''\mathscr{W}. \tag{12}$$

La aplicación $\mathscr{U}$ es lineal, ya que si

$$y = y' + y'' \quad (y' \in \mathfrak{A} \quad \text{e} \quad y'' \in \mathfrak{A}^{\perp}),$$

de (7), (9) y (11) resulta:

$$(\alpha x + \beta y)\mathscr{U} = (\alpha x' + \beta y')\mathscr{V} + (\alpha x'' + \beta y'')\mathscr{W} = \alpha(x\mathscr{U}) + \beta(y\mathscr{U}).$$

La aplicación $\mathscr{U}$ es unitaria, ya que debido a (8), (10) y (11) se tiene

$$(x\mathscr{U}, x\mathscr{U}) = (x'\mathscr{V} + x''\mathscr{W}, x'\mathscr{V} + x''\mathscr{W}) =$$
$$= (x'\mathscr{V}, x'\mathscr{V}) + (x''\mathscr{W}, x''\mathscr{W}) = (x', x') + (x'' + x'') = (x, x).$$

Para todo x de $\mathfrak{L}$ es válida la relación

$$x\mathscr{A}\mathscr{U} = x\mathscr{B}.$$

Efectivamente, $x\mathscr{A}$ pertenece a $\mathfrak{A}$; luego, el vector x'' de la descomposición (11) es igual a cero y, por consiguiente,

$$x\mathscr{A}\mathscr{U} = x\mathscr{A}\mathscr{V}.$$

Teniendo en cuenta (6), esto da $x\mathscr{A}\mathscr{U} = x\mathscr{B}$. De aquí resulta $\mathscr{A}\mathscr{U} = \mathscr{B}$ y el lema queda demostrado.

TEOREMA 1. *Toda aplicación lineal $\mathscr{A}$ de un espacio unitario $\mathfrak{L}$ admite una descomposición polar*

$$\mathscr{A} = \mathscr{D}\mathscr{U}, \tag{13}$$

donde $\mathscr{D}$ es una aplicación simétrica no negativa y $\mathscr{U}$ es una aplicación unitaria del espacio $\mathfrak{L}$. La aplicación $\mathscr{D}$ se determina unívocamente; si $\mathscr{A}$ es una aplicación regular, la aplicación $\mathscr{U}$ también se determina unívocamente.

La aplicación $\mathcal{A}\mathcal{A}^*$ es, según el p. 19.6, simétrcia y no negativa. Indiquemos por $\mathcal{D}$ la raíz cuadrada simétrica no negativa de $\mathcal{A}\mathcal{A}^*$. Es decir,

$$\mathcal{D}^2 = \mathcal{A}\mathcal{A}^*.$$

Tenemos cualquiera que sea el vector x

$$(x\mathcal{A},\ x\mathcal{A}) = (x,\ x\mathcal{A}\mathcal{A}^*) = (x,\ x\mathcal{D}^2) = (x\mathcal{D},\ x\mathcal{D}),$$

es decir, las aplicaciónes $\mathcal{A}$ y $\mathcal{D}$ alteran igualmente las longitudes de los vectores. Basándonos en el lema, deducimos de aquí que existe una aplicación unitaria $\mathcal{U}$ tal que

$$\mathcal{D}\mathcal{U} = \mathcal{A}.$$

Con esto queda demostrada la existencia de la descomposición polar. Resta examinar su unicidad. De (13) resulta $\mathcal{A}^* = \mathcal{U}^*\mathcal{D} = \mathcal{U}^{-1}\mathcal{D}$ y $\mathcal{A}\mathcal{A}^* = \mathcal{D}\mathcal{U}\mathcal{U}^{-1}\mathcal{D} = \mathcal{D}^2$. Por consiguiente, la aplicación $\mathcal{D}$ es no negativa y simétrica y su cuadrado es igual a la aplicación $\mathcal{A}\mathcal{A}^*$. Según el teorema 7 del p. 19.6, estas condiciones determinan unívocamente la aplicación $\mathcal{D}$. Si $\mathcal{A}$ es una aplicación regular, también $\mathcal{D}$ es regular y de (13) resulta entonces que $\mathcal{U} = \mathcal{D}^{-1}\mathcal{A}$, es decir, la aplicación $\mathcal{U}$ también se determina unívocamente.

El significado geométrico del teorema 1 es muy sencillo. Indica precisamente que la acción de toda aplicación lineal del espacio $\mathfrak{L}$ puede ser representada de la forma siguiente: primero el espacio $\mathfrak{L}$ se dilata en n direcciones recíprocamente ortogonales con un coeficiente de dilatación concreto, real y no negativo, en cada una de las direcciones y después gira alrededor del origen de coordenadas [1]. Si la aplicación es regular, todos los coeficientes de dilatación son estrictamente positivos. En el caso de una aplicación singular algunos de los coeficientes resultan iguales a cero y en lugar de la dilatación en las direcciones correspondientes tiene lugar la proyección del espacio.

Observemos también que en la demostración de la existencia de la descomposición polar nos hemos basado en el producto $\mathcal{A}\mathcal{A}^*$. Si en lugar de él tomamos el producto $\mathcal{A}^*\mathcal{A}$, obtendremos la descomposición de tipo

$$\mathcal{A} = \mathcal{U}\mathcal{D}_1,$$

donde $\mathcal{U}$ es unitaria y $\mathcal{D}_1$ es una aplicación simétrica no negativa.

Tomemos en el espacio $\mathfrak{L}$ un sistema ortonormal de coordenadas. A las aplicaciones unitarias les corresponden entonces matrices unitarias y a las aplicaciones simétricas les corresponden matrices hermitianas y el teorema 1 se convierte en la proposición siguiente: *toda matriz cuadrada puede ser representada en la forma de un producto de una matriz hermitiana y otra unitaria.*

[1] El giro se entiende en el sentido de una aplicación unitaria.

Suponiendo que el campo principal es el cuerpo de los números reales, obtenemos que *toda matriz cuadrada real puede ser representada como el producto de una matriz simétrica real y otra ortogonal real.*

En el teorema 1 se afirma que $\mathcal{D}$ es una aplicación simétrica *no negativa*. De acuerdo con esto en las dos últimas proposiciones a las palabras *hermitiana* y *simétrica* se puede agregar: *de valores propios no negativos.*

20.3. Aplicación de Cayley. Comparando las propiedades de las aplicaciones unitarias con las propiedades de las aplicaciones simétricas, podemos notar que ambas clases de aplicaciones están ligadas estrechamente. En forma explícita esta relación queda expresada en las así llamadas fórmulas de Cayley.

TEOREMA 2 (aplicación de Cayley). *Si $\mathcal{A}$ es una aplicación simétrica de un espacio unitario complejo, las aplicaciones $\mathcal{A} \pm i\mathcal{E}$ son invertibles; la aplicación $\mathcal{U}$ definida mediante la fórmula*

$$\mathcal{U} = (\mathcal{A} - i\mathcal{E})(\mathcal{A} + i\mathcal{E})^{-1} \tag{14}$$

es unitaria, no tiene valores propios iguales a la unidad y, además, $\mathcal{A}$ se expresa mediante $\mathcal{U}$ por la fórmula

$$\mathcal{A} = -i(\mathcal{U} + \mathcal{E})(\mathcal{U} - \mathcal{E})^{-1}. \tag{15}$$

Recíprocamente, si $\mathcal{U}$ es una aplicación unitaria que no tiene valores propios iguales a la unidad, la aplicación $\mathcal{U} - \mathcal{E}$ es invertible, la aplicación $\mathcal{A}$ calculada mediante la fórmula (15) *es simétrica y $\mathcal{U}$ se expresa mediante $\mathcal{A}$ en la forma* (14).

DEMOSTRACIÓN. Sea $\mathcal{A}$ una aplicación simétrica de un espacio unitario. Los números $\pm i$ no pueden ser valores propios de la aplicación $\mathcal{A}$, ya que todos los valores propios de las aplicaciones simétricas son reales (p. 19.4). Esto significa que las aplicaciones $\mathcal{A} \pm i\mathcal{E}$ son regulares. Puesto que las aplicaciones $\mathcal{A} + i\mathcal{E}$ y $\mathcal{A} - i\mathcal{E}$ conmutan, resulta que ellas conmutan también con las aplicaciones $(\mathcal{A} + i\mathcal{E})^{-1}$ y $(\mathcal{A} - i\mathcal{E})^{-1}$. Para la aplicación $\mathcal{U}$ definida mediante la fórmula (14) existe la conjugada que es igual a

$$\mathcal{U}^* = (\mathcal{A} + i\mathcal{E})^{*-1}(\mathcal{A} - i\mathcal{E})^* = (\mathcal{A} - i\mathcal{E})^{-1}(\mathcal{A} + i\mathcal{E}).$$

Tenemos de aquí

$$\mathcal{U}\mathcal{U}^* = (\mathcal{A} - i\mathcal{E})(\mathcal{A} + i\mathcal{E})^{-1}(\mathcal{A} - i\mathcal{E})^{-1}(\mathcal{A} + i\mathcal{E}) =$$
$$= (\mathcal{A} - i\mathcal{E})(\mathcal{A} + i\mathcal{E})^{-1}(\mathcal{A} + i\mathcal{E})(\mathcal{A} - i\mathcal{E})^{-1} = \mathcal{E},$$

es decir, $\mathcal{U}$ es unitaria. Probemos que $\mathcal{U} - \mathcal{E}$ es invertible. Para ello restemos de ambos miembros de la igualdad (14) la aplicación $\mathcal{E}$ y multipliquemos los resultados por $\mathcal{A} + i\mathcal{E}$. Después de efectuar transformaciones evidentes, tendremos

$$(\mathcal{U} - \mathcal{E})(\mathcal{A} + i\mathcal{E}) = -2i\mathcal{E}, \tag{16}$$

es decir,

$$(\mathcal{U}-\mathcal{E})^{-1}=-\frac{1}{2i}(\mathcal{A}+i\mathcal{E}).$$

Por consiguiente, $\mathcal{U}$ no tiene valores propios iguales a la unidad. Además, de (16) resulta:

$$(\mathcal{U}-\mathcal{E})\mathcal{A}=-2i\mathcal{E}-l(\mathcal{U}-\mathcal{E})=-i(\mathcal{U}+\mathcal{E}),$$

es decir,

$$\mathcal{A}=-i(\mathcal{U}+\mathcal{E})(\mathcal{U}-\mathcal{E})^{-1}.$$

Hemos demostrado la primera parte del teorema. La demostración de la reciproca es análoga totalmente a la demostración realizada.

Las fórmulas de Cayley establecen una correspondencia biyectiva entre todas las aplicaciones simétricas de un espacio unitario $\mathfrak{L}$ y aquellas aplicaciones unitarias $\mathcal{U}$ del mismo para los cuales 1 no es valor propio. Análogamente, las fórmulas

$$\mathcal{U}=(i\mathcal{E}+\mathcal{A})(i\mathcal{E}-\mathcal{A})^{-1}, \tag{14'}$$

$$\mathcal{A}=i(\mathcal{U}-\mathcal{E})(\mathcal{U}+\mathcal{E})^{-1} \tag{15'}$$

ofrecen una correspondencia biyectiva entre las aplicaciones simétricas del espacio $\mathfrak{L}$ y aquellas aplicaciones unitarias $\mathcal{U}$ para las cuales -1 no es valor propio.

Las aplicaciones (14) y (15), así como las aplicaciones (14') y (15'), son posibles gracias a que en el campo principal existe el número i. Si el campo principal es real, las fórmulas indicadas no son válidas. Sin embargo, es fácil modificar estas fórmulas de manera que sean válidas para cualquier campo. Tiene lugar el teorema siguiente:

TEOREMA 3. *Sea $\mathcal{A}$ una aplicación antisimétrica de un espacio unitario $\mathfrak{L}$. Entonces las aplicaciones $\mathcal{A}\pm\mathcal{E}$ son invertibles, la aplicación*

$$\mathcal{U}=(\mathcal{A}-\mathcal{E})(\mathcal{A}+\mathcal{E})^{-1} \tag{17}$$

es unitaria, no tiene valores propios iguales a la unidad y, además,

$$\mathcal{A}=-(\mathcal{U}+\mathcal{E})(\mathcal{U}-\mathcal{E})^{-1}. \tag{18}$$

Recíprocamente, si $\mathcal{U}$ es una aplicación unitaria y la unidad no es valor propio de la misma, la aplicación $\mathcal{A}$ definida mediante la fórmula (18) *es antisimétrica y $\mathcal{U}$ se expresa mediante $\mathcal{A}$ en la forma* (17).

En efecto, si $\mathcal{A}$ es antisimétrica, los números ± 1 no pueden ser sus valores propios, ya que todos los valores propios de las aplicaciones antisimétricas o bien son iguales a cero o bien son imaginarios puros (p. 19.5). Por esto las aplicaciones $\mathcal{A}\pm\mathcal{E}$ son regulares. Puesto que $(\mathcal{A}+\mathcal{E})(\mathcal{A}-\mathcal{E})=(\mathcal{A}-\mathcal{E})(\mathcal{A}+\mathcal{E})$ tenemos $(\mathcal{A}-\mathcal{E})(\mathcal{A}+\mathcal{E})^{-1}=(\mathcal{A}+\mathcal{E})^{-1}(\mathcal{A}-\mathcal{E})$. De (17) resulta

$$\mathcal{U}^*=(\mathcal{A}^*+\mathcal{E})^{-1}(\mathcal{A}^*-\mathcal{E})=(-\mathcal{A}+\mathcal{E})^{-1}(-\mathcal{A}-\mathcal{E})=$$
$$=(\mathcal{A}-\mathcal{E})^{-1}(\mathcal{A}+\mathcal{E}),$$

de donde

$$\mathcal{U}\mathcal{U}^* = (\mathcal{A}-\mathcal{E})(\mathcal{A}+\mathcal{E})^{-1}(\mathcal{A}-\mathcal{E})^{-1}(\mathcal{A}+\mathcal{E}) = \mathcal{E},$$

es decir, la aplicación $\mathcal{U}$ es unitaria. Los razonamientos ulteriores son totalmente análogos a los realizados en la demostración del teorema 2 y, por ello, los omitimos.

Para concluir, observemos que los resultados de los últimos parágrafos descubren cierta semejanza entre las propiedades de las aplicaciones lineales de los espacios unitarios y las propiedades de los números complejos. Convengamos en aceptar que las aplicaciones lineales son, en cierto sentido, análogas a los números complejos y que las aplicaciones conjugadas son análogas a los números conjugados. Entonces las aplicaciones simétricas, que se caracterizan por la condición $\mathcal{A}^* = \mathcal{A}$, serán análogas a los números complejos que satisfacen la relación $\bar{z} = z$, es decir, a los números reales; las aplicaciones antisimétricas, que se caracterizan por la condición $\mathcal{A}^* = -\mathcal{A}$, serán análogas a los números complejos que satisfacen a la relación $\bar{z} = -z$, es decir, a los números imaginarios puros; las aplicaciones unitarias con la propiedad $\mathcal{U}\mathcal{U}^* = \mathcal{E}$ serán análogas a los números complejos z para los cuales $z\bar{z} = 1$, es decir, $|z| = 1$. La descomposición $\mathcal{A} = \mathcal{B} + i\mathcal{C}$ del punto 20.1 corresponderá a la representación de un número complejo z en la forma cartesiana $z = x + iy$ y la descomposición polar $\mathcal{A} = \mathcal{D}\mathcal{U}$ corresponderá a la representación de un número complejo en la forma trigonométrica $z = \rho(\cos\varphi + i\,\mathrm{sen}\,\varphi)$, etc.

20.4. Descomposición espectral. Desde el punto de vista geométrico uno de los tipos más sencillos de las aplicaciones lineales es la proyección de los vectores sobre un subespacio. Algunas de las propiedades de estas aplicaciones proyectivas serán ahora consideradas.

Sea $\mathfrak{A}$ un subespacio lineal de un espacio unitario $\mathfrak{L}$. El conjunto de vectores ortogonales a $\mathfrak{A}$ es el subespacio ortogonal $\mathfrak{A}^\perp$ y $\mathfrak{L}$ es la suma directa de $\mathfrak{A}$ y $\mathfrak{A}^\perp$. Luego, todo vector a de $\mathfrak{L}$ puede ser representado unívocamente en la forma

$$a = a' + a'' \quad (a' \in \mathfrak{A} \quad \text{y} \quad a'' \in \mathfrak{A}^\perp). \tag{19}$$

El vector a' se llama *proyección* del vector a sobre el subespacio $\mathfrak{A}$. Poniendo en correspondencia a todo vector su proyección sobre $\mathfrak{A}$, obtenemos una aplicación del espacio $\mathfrak{L}$ que se llama *proyectiva* y se indica por $\mathcal{P}_{\mathfrak{A}}$. Es decir, se toma por definición

$$a\mathcal{P}_{\mathfrak{A}} = a'.$$

A veces, para abreviar la notación, omitiremos el índice $\mathfrak{A}$ y en lugar de $\mathcal{P}_{\mathfrak{A}}$ escribiremos $\mathcal{P}$.

Las aplicaciones proyectivas son lineales, ya que si para un vector a tiene lugar la descomposición (19) y para otro vector b la descomposición

$$b = b' + b'' \quad (b' \in \mathfrak{A} \quad \text{y} \quad b'' \in \mathfrak{A}^\perp),$$

resulta

$$\alpha a + \beta b = (\alpha a' + \beta b') + (\alpha a'' + \beta b''),$$

es decir, $(\alpha a+\beta b)\mathcal{P}=\alpha a'+\beta b'$, de donde

$$(\alpha a+\beta b)\mathcal{P}=\alpha\cdot a\mathcal{P}+\beta\cdot b\mathcal{P}.$$

Las aplicaciones proyectivas son aplicaciones simétricas no negativas. Efectivamente, de la descomposición (19) tenemos

$$(a\mathcal{P},\ a)=(a',\ a'+a'')=(a',\ a')\geqslant 0,$$

es decir, $\mathcal{P}$ es no negativa. Por otra parte, si

$$b=b'+b''\quad (b'\in\mathfrak{A}\ \text{ y }\ b''\in\mathfrak{A}^{\perp}),$$

tenemos

$$(a\mathcal{P},\ b)=(a',\ b'+b'')=(a',\ b')=(a'+a'',\ b')=(a,\ b\mathcal{P}),$$

es decir, la aplicación $\mathcal{P}$ es simétrica.

De la fórmula (19) se desprende también otra propiedad importante de las aplicaciones proyectivas. Es evidente que para cualquier a tenemos

$$a\mathcal{P}^2=a'\mathcal{P}=a'=a\mathcal{P},$$

de donde

$$\mathcal{P}^2=\mathcal{P}. \tag{20}$$

Las aplicaciones que coinciden con su cuadrado se llaman *idempotentes*. Por consiguiente, de (20) resulta que *todas las aplicaciones proyectivas son idempotentes.*

Recíprocamente, las propiedades de simetría y de idempotencia caracterizan plenamente las aplicaciones proyectivas: *toda aplicación simétrica idempotente $\mathcal{P}$ es la proyección sobre el dominio de valores de $\mathcal{P}$.*

Sea $\mathfrak{A}$ el dominio de valores de $\mathcal{P}$. Para todo a es válida la descomposición

$$a=a\mathcal{P}+a(\mathcal{E}-\mathcal{P}). \tag{21}$$

El sumando $a\mathcal{P}$ pertenece, por definición, a $\mathfrak{A}$. El segundo sumando es ortogonal a $\mathfrak{A}$, ya que todo vector de $\mathfrak{A}$ es de la forma $x\mathcal{P}$, donde x es un vector de $\mathfrak{L}$, y debido a que la aplicación $\mathcal{P}$ es simétrica e idempotente tenemos

$$(x\mathcal{P},\ a(\mathcal{E}-\mathcal{P}))=(x,\ a(\mathcal{E}-\mathcal{P})\mathcal{P})=(x,\ a\mathcal{P}-a\mathcal{P})=0.$$

Por lo tanto, la descomposición (21) muestra que $a\mathcal{P}$ es la proyección del vector a sobre $\mathfrak{A}$ que es lo que se quería demostrar.

Hallemos la forma más sencilla de la matriz de una aplicación proyectiva $\mathcal{P}_{\mathfrak{A}}$. Tomemos en $\mathfrak{A}$ y en $\mathfrak{A}^{\perp}$ unas bases ortonormales $e_1,\ \ldots,\ e_m$ y $e_{m+1},\ \ldots,\ e_n$. El sistema unido $e_1,\ \ldots,\ e_m,\ e_{m+1},\ \ldots,\ e_n$ será una base ortonormal del espacio $\mathfrak{L}$. Las igualdades

$$e_j\mathcal{P}=e_j\quad\text{y}\quad e_\alpha\mathcal{P}=0\quad (j=1,\ \ldots,\ m;\ \alpha=m+1,\ \ldots,\ n)$$

muestran que la matriz de la aplicación $\mathcal{P}_{\mathfrak{A}}$ en esta base es de la forma

$$P=\begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix}, \tag{22}$$

Recíprocamente, si en un sistema ortonormal de coordenadas la matriz de una aplicación lineal $\mathcal{P}$ se reduce a la forma (22), es evidente que $\mathcal{P}$ es una aplicación proyectiva.

Sean $\mathcal{P}$ y Q las operaciones de proyección de un espacio $\mathfrak{L}$ sobre unos subespacios $\mathfrak{A}$ y $\mathfrak{B}$. Surge la pregunta: ¿cómo la posición de los subespacios $\mathfrak{A}$ y $\mathfrak{B}$ en $\mathfrak{L}$ influye en las propiedades de las aplicaciones $\mathcal{P}$ y Q? Por ejemplo, ¿qué puede decirse sobre $\mathcal{P}$ y Q, si $\mathfrak{A}$ y $\mathfrak{B}$ son ortogonales o si $\mathfrak{A}$ pertenece a $\mathfrak{B}$, etc.? Para responder a estas preguntas introduciremos, ante todo, la definición siguiente: *dos aplicaciones proyectivas $\mathcal{P}$ y Q se llaman ortogonales, si $\mathcal{P}Q=\mathcal{O}$.* Puesto que las aplicaciones proyectivas son simétricas, tenemos de aquí

$$Q\mathcal{P}=Q^*\mathcal{P}^*=(\mathcal{P}Q)^*=\mathcal{O},$$

es decir, si $\mathcal{P}$ es ortogonal a Q, Q es ortogonal a $\mathcal{P}$.

Para que unas aplicaciones proyectivas $\mathcal{P}$ y Q sean ortogonales es necesario y suficiente que sean ortogonales los subespacios correspondientes $\mathfrak{A}$ y $\mathfrak{B}$.

Sea $\mathcal{P}Q=\mathcal{O}$; entonces para $a\in\mathfrak{A}$ y $b\in\mathfrak{B}$ tenemos

$$(a,\ b)=(a\mathcal{P},\ bQ)=(a\mathcal{P}Q,\ b)=0,$$

es decir, $\mathfrak{A}$ es ortogonal a $\mathfrak{B}$. Recíprocamente, si $\mathfrak{A}$ es ortogonal a $\mathfrak{B}$, tenemos para cualquier vector x de $\mathfrak{L}$

$$x\mathcal{P}\in\mathfrak{A},\ x(\mathcal{P}Q)=(x\mathcal{P})Q=o\quad\text{y}\quad\mathcal{P}Q=\mathcal{O}$$

que es lo que se quería demostrar.

Para el estudio detallado de las propiedades de las aplicaciones lineales suelen emplearse las representaciones matriciales de las mismas. Pero si la representación matricial es, por cualquier razón, incómoda, se trata de expresar la aplicación lineal dada mediante aplicaciones de carácter más simple. En el caso de las aplicaciones normales, para estas aplicaciones elementales se pueden tomar las aplicaciones proyectivas.

Una descomposición de tipo

$$\mathcal{A}=\alpha_1\mathcal{P}_1+\alpha_2\mathcal{P}_2+\ldots\alpha_s\mathcal{P}_s \tag{23}$$

se llama descomposición espectral de la aplicación $\mathcal{A}$, si

a) *los números $\alpha_1,\ \ldots,\ \alpha_s$ son diferentes;*

b) $\mathcal{P}_j^2=\mathcal{P}_j\neq\mathcal{O}\quad(j=1,\ 2,\ \ldots,\ s)$;

c) $\mathcal{P}_j^*=\mathcal{P}_j\quad(j=1,\ 2,\ \ldots,\ s)$;

d) $\mathcal{P}_j\mathcal{P}_k=\mathcal{O}\quad(j\neq k;\ j,\ k=1,\ 2,\ \ldots,\ s)$;

e) $\mathcal{P}_1+\mathcal{P}_2+\ldots+\mathcal{P}_s=\mathcal{E}$.

Las condiciones b), c) y d) significan que $\mathcal{P}_1,\ \ldots,\ \mathcal{P}_s$ son aplicaciones proyectivas recíprocamente ortogonales.

Está claro que *sólo las aplicaciones normales admiten la descomposición espectral*. Efectivamente, de (23) se deduce:

$$\mathcal{A}^*=\bar\alpha_1\mathcal{P}_1+\bar\alpha_2\mathcal{P}_2+\ldots+\bar\alpha_s\mathcal{P}_s,$$
$$\mathcal{A}\mathcal{A}^*=\Big(\sum\alpha_j\mathcal{P}_j\Big)\Big(\sum\bar\alpha_k\mathcal{P}_k\Big)=\sum\bar\alpha_j\alpha_j\mathcal{P}_j=\mathcal{A}^*\mathcal{A}.$$

Recíprocamente, *toda aplicación normal de un espacio unitario complejo admite una descomposición espectral.*

Sea $\mathcal{A}$ una aplicación normal de un espacio unitario complejo $\mathfrak{L}$. Hemos visto que en $\mathfrak{L}$ existe una base ortonormal $e_1,\ e_2,\ \ldots,\ e_n$ formada por los vectores propios de la aplicación $\mathcal{A}$. Ordenemos estos vectores de manera que aquellos que corresponden a valores propios iguales se encuentren al lado. Supongamos, por ejemplo, que $e_1,\ \ldots,\ e_{m_1}$ corresponden al valor propio α_1, que $e_{m_1+1},\ \ldots,\ e_{m_2}$ corresponden al valor propio α_2, etc. Indiquemos por $\mathfrak{L}_j$ el subespacio tendido sobre los vectores $e_{m_{j-1}+1},\ \ldots,\ e_{m_j}$ correspondientes al

valor propio α_i ($i=1, 2, \ldots, s$). Puesto que todos los vectores coordenados de $\mathfrak{L}_i$ son vectores propios correspondientes a un mismo valor propio α_i, todos los vectores de $\mathfrak{L}_i$ serán vectores propios con valores propios iguales a α_i. Tenemos

$$\mathfrak{L}=\mathfrak{L}_1+\mathfrak{L}_2+\ldots+\mathfrak{L}_s, \tag{24}$$

donde los subespacios $\mathfrak{L}_1, \ldots, \mathfrak{L}_s$ son recíprocamente ortogonales. Indiquemos por $\mathcal{P}_i$ la aplicación de proyección sobre el subespacio $\mathfrak{L}_i$. Pero la ortogonalidad de los subespacios $\mathfrak{L}_i$ implica la ortogonalidad de las correspondientes aplicaciones proyectivas. Además, de la igualdad (24) se deduce que $\mathcal{P}_1+\mathcal{P}_2+\ldots+\mathcal{P}_s=\mathcal{P}$, donde $\mathcal{P}$ es la proyección sobre $\mathfrak{L}$. Como la proyección sobre $\mathfrak{L}$ es la aplicación idéntica, tenemos

$$\mathcal{E}=\mathcal{P}_1+\mathcal{P}_2+\ldots+\mathcal{P}_s. \tag{25}$$

Vemos, por consiguiente, que las aplicaciones $\mathcal{P}_1, \ldots, \mathcal{P}_s$ poseen las propiedades de b) a e). Probemos finalmente que

$$\mathcal{A}=\alpha_1\mathcal{P}_1+\alpha_2\mathcal{P}_2+\ldots+\alpha_s\mathcal{P}_s.$$

Sea a un vector cualquiera de $\mathfrak{L}$. De la igualdad (25) resulta

$$a=a\mathcal{P}_1+a\mathcal{P}_2+\ldots+a\mathcal{P}_s. \tag{26}$$

El vector $a\mathcal{P}_i$ pertenece a $\mathfrak{L}_i$ y todos los vectores de $\mathfrak{L}_i$ son vectores propios correspondientes al valor propio α_i; por ello $a\mathcal{P}_i\mathcal{A}=\alpha_i\cdot a\mathcal{P}_i$. Multiplicando (26) por $\mathcal{A}$, obtenemos

$$a\mathcal{A}=\alpha_1 a\mathcal{P}_1+\ldots+\alpha_s a\mathcal{P}_s=a\,(\alpha_1\mathcal{P}_1+\ldots+\alpha_s\mathcal{P}_s),$$

de donde $\mathcal{A}=\alpha_1\mathcal{P}_1+\ldots+\alpha_s\mathcal{P}_s$ que es lo que se quería demostrar.

Si

$$\mathcal{A}=\alpha_1\mathcal{P}_1+\alpha_2\mathcal{P}_2+\ldots+\alpha_s\mathcal{P}_s \tag{27}$$

es una descomposición espectral de una aplicación $\mathcal{A}$, entonces $\alpha_1, \ldots, \alpha_s$ es el conjunto de todos los valores propios de esta aplicación.

En efecto, cualquier aplicación $\mathcal{P}_j$ es, por hipótesis, diferente de la aplicación $\mathcal{O}$. Por consiguiente, existe un vector a tal que $a\mathcal{P}_j\neq o$. Pero como las aplicaciones $\mathcal{P}_1, \ldots, \mathcal{P}_s$ son ortogonales, obtenemos entonces

$$a\mathcal{P}_j\mathcal{A}=\alpha_1 a\mathcal{P}_j\mathcal{P}_1+\alpha_2 a\mathcal{P}_j\mathcal{P}_2+\ldots+\alpha_s a\mathcal{P}_j\mathcal{P}_s=\alpha_j a\mathcal{P}_j,$$

es decir, $a\mathcal{P}_j$ es un vector propio correspondiente al valor propio α_j.

Recíprocamente, sea a un vector propio no nulo de la aplicación $\mathcal{A}$ correspondiente al valor propio β. De la propiedad e) se deduce que

$$a=a\mathcal{P}_1+a\mathcal{P}_2+\ldots+a\mathcal{P}_s. \tag{28}$$

La condición (27) da:

$$a\mathcal{A}=\alpha_1 a\mathcal{P}_1+\alpha_2 a\mathcal{P}_2+\ldots+\alpha_s a\mathcal{P}_s.$$

Puesto que $a\mathcal{A}=\beta a$, tenemos

$$\alpha_1 a\mathcal{P}_1+\ldots+\alpha_s a\mathcal{P}_s=\beta a\mathcal{P}_1+\ldots+\beta a\mathcal{P}_s.$$

Multiplicando esta relación por $\mathcal{P}_j$ y empleando las propiedades c) y d) encontramos

$$\alpha_j a\mathcal{P}_j=\beta a\mathcal{P}_j \quad\text{y}\quad (\alpha_j-\beta)\,a\mathcal{P}_j=o\ (j=1, \ldots, s).$$

El vector a es diferente de cero y por ello al menos un sumando $a\mathcal{P}_j$ de (28) es también diferente de cero. Pero de la igualdad $(\alpha_j-\beta)\,a\mathcal{P}_j=o$ resulta entonces $\beta=\alpha_j$ que es lo que se quería demostrar.

Si $\mathcal{A}=\alpha_1\mathcal{P}_1+\ldots+\alpha_s\mathcal{P}_s$ *es una descomposición espectral de una aplicación* $\mathcal{A}$ *y si* $f(\lambda)=\beta_0+\beta_1\lambda+\ldots+\beta_m\lambda^m$ *es un polinomio cualquiera, se tiene*

$$f(\mathcal{A})=f(\alpha_1)\mathcal{P}_1+\ldots+f(\alpha_s)\mathcal{P}_s. \tag{29}$$

Tenemos

$$\mathcal{E}=\mathcal{P}_1+\ldots+\mathcal{P}_s,$$
$$\mathcal{A}=\alpha_1\mathcal{P}_1+\ldots+\alpha_s\mathcal{P}_s,$$
$$\mathcal{A}^2=\sum\alpha_j\mathcal{P}_j\cdot\sum\alpha_j\mathcal{P}_j=\sum\alpha_j\alpha_k\mathcal{P}_j\mathcal{P}_k=\alpha_1^2\mathcal{P}_1+\ldots+\alpha_s^2\mathcal{P}_s,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\mathcal{A}^m=\sum\alpha_j^{m-1}\mathcal{P}_j\cdot\sum\alpha_j\mathcal{P}_j=\sum\alpha_j^{m-1}\alpha_k\mathcal{P}_j\mathcal{P}_k=\alpha_1^m\mathcal{P}_1+\ldots+\alpha_s^m\mathcal{P}_s.$$

Multiplicando estas igualdades por los números $\beta_0, \beta_1, \ldots, \beta_m$, respectivamente, y sumándolas obtenemos la relación requerida.

Si

$$\mathcal{A}=\alpha_1\mathcal{P}_1+\alpha_2\mathcal{P}_2+\ldots+\alpha_s\mathcal{P}_s$$

es una descomposición espectral de una aplicación $\mathcal{A}$, *se* tiene

$$\mathcal{P}_i=\frac{(\mathcal{A}-\alpha_1\mathcal{E})\ldots(\mathcal{A}-\alpha_{i-1}\mathcal{E})(\mathcal{A}-\alpha_{i+1}\mathcal{E})\ldots(\mathcal{A}-\alpha_s\mathcal{E})}{(\alpha_i-\alpha_1)\ldots(\alpha_i-\alpha_{i-1})(\alpha_i-\alpha_{i+1})\ldots(\alpha_i-\alpha_s)}.$$

En efecto, consideremos el polinomio

$$\varphi_i(\lambda)=\frac{(\lambda-\alpha_1)\ldots(\lambda-\alpha_{i-1})(\lambda-\alpha_{i+1})\ldots(\lambda-\alpha_s)}{(\alpha_i-\alpha_1)\ldots(\alpha_i-\alpha_{i-1})(\alpha_i-\alpha_{i+1})\ldots(\alpha_i-\alpha_s)}.$$

De acuerdo con (29) tenemos

$$\varphi_i(\mathcal{A})=\varphi_i(\alpha_1)\mathcal{P}_1+\varphi_i(\alpha_2)\mathcal{P}_2+\ldots+\varphi_i(\alpha_s)\mathcal{P}_s.$$

Pero $\varphi_i(\alpha_j)=0$ $(i\neq j)$ y $\varphi_i(\alpha_i)=1$; por consiguiente, $\varphi_i(\mathcal{A})=\mathcal{P}_i$.

La última propiedad significa que *toda aplicación normal de un espacio unitario admite una descomposición espectral, y sólo una.*

Efectivamente, ya hemos demostrado la existencia de la descomposición y por ello sólo se trata de demostrar su unicidad. Los coeficientes $\alpha_1, \ldots, \alpha_s$ de la descomposición espectral son los valores propios de la aplicación $\mathcal{A}$ y, por consiguiente, se determinan unívocamente por la aplicación $\mathcal{A}$. Pero si conocemos estos coeficientes, conocemos al mismo tiempo los polinomios $\varphi_i(\lambda)$ y, por lo tanto, conocemos las aplicaciones proyectivas $\mathcal{P}_i$ que es lo que se quería demostrar.

Indiquemos algunas otras propiedades de las descomposiciones espectrales. El conjunto de los coeficientes $\alpha_1, \ldots, \alpha_s$ de la descomposición espectral se denomina *espectro* de la misma. Se llama *espectro de una aplicación lineal* el espectro de su descomposición espectral. Hemos demostrado anteriormente que el espectro de una aplicación normal coincide con el conjunto de sus valores propios.

Para que una aplicación normal sea simétrica, respectivamente antisimétrica o unitaria, es necesario y suficiente que su espectro sea real, respectivamente imaginario puro o formado por números de módulo unidad.

Sea $\mathcal{A}=\alpha_1\mathcal{P}_1+\alpha_2\mathcal{P}_2+\ldots+\alpha_s\mathcal{P}_s$ la descomposición espectral de la aplicación $\mathcal{A}$. Tenemos entonces que $\mathcal{A}^*=\bar{\alpha}_1\mathcal{P}_1+\bar{\alpha}_2\mathcal{P}_2+\ldots+\bar{\alpha}_s\mathcal{P}_s$ es la descomposición espectral de la aplicación conjugada. Puesto que toda aplicación normal admite sólo una descomposición espectral, resulta que la condición $\mathcal{A}=\mathcal{A}^*$ equivale a las igualdades $\alpha_i=\bar{\alpha}_i$ $(i=1, \ldots, s)$, es decir, equivale a que el espectro sea real. Hemos demostrado la primera afirmación. Las otras dos se demuestran análogamente.

Ejemplos y problemas

1. Descompóngase en parte simétrica y antisimétrica las matrices

$$\begin{bmatrix} 2 & 7 & 0 \\ 6 & 2 & 2 \\ 2 & 0 & 5 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 0 \\ 4 & 0 & 0 \\ 0 & 0 & 9 \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} -1 & 2 & -2 \\ 4 & -2 & -4 \\ 4 & 4 & 2 \end{bmatrix}.$$

y hállense las descomposiciones polares de las mismas.

2. Sea $\mathfrak{L}$ un espacio euclídeo de dos dimensiones y sea $\mathcal{A}$ una aplicación lineal del mismo que tiene dos vectores propios unitarios no ortogonales a_1 y a_2 correspondientes a los valores propios α_1 y α_2 ($\alpha_1 \neq \alpha_2$). Hállese la descomposición polar de la aplicación $\mathcal{A}$, si $(a_1, a_2) = \cos\varphi$, donde φ es un valor dado.

3. Indíquese el enunciado matricial de la aplicación de Cayley (teoremas 2 y 3).

4. Demuéstrese que toda matriz ortogonal U de orden tres, que no contiene la unidad entre sus valores propios, es de la forma

$$U = \frac{1}{1+\alpha^2+\beta^2+\gamma^2} \begin{bmatrix} \alpha^2+\beta^2-\gamma^2-1 & 2(\alpha+\beta\gamma) & 2(\beta-\alpha\gamma) \\ 2(-\alpha+\beta\gamma) & \alpha^2-\beta^2+\gamma^2-1 & 2(\gamma+\alpha\beta) \\ 2(-\beta-\alpha\gamma) & 2(-\gamma+\alpha\beta) & -\alpha^2+\beta^2+\gamma^2-1 \end{bmatrix}.$$

5. Las matrices cuadradas complejas de orden n forman, respecto a las operaciones de adición y de multiplicación por número, un espacio lineal $\mathfrak{L}$ de dimensión n^2. Las matrices E_{ij}, en la i-ésima fila y j-ésima columna de las cuales aparece la unidad, mientras que en las demás posiciones aparece el cero, constituyen una base del espacio $\mathfrak{L}$. Hagamos el espacio $\mathfrak{L}$ unitario, aceptando que las matrices E_{ij} ofrecen una base ortonormal en $\mathfrak{L}$, es decir, aceptando que el producto escalar de dos matrices A y B se calcula mediante la fórmula $(A, B) = \sum\sum \alpha_{ij}\overline{\beta}_{ij}$, donde α_{ij} son los elementos de la matriz A y β_{ij} son los elementos de la matriz B. La multiplicación de todas las matrices de $\mathfrak{L}$ por una matriz cualquiera X a la derecha ofrece una aplicación lineal del espacio $\mathfrak{L}$. Demuéstrese que:

a) las matrices unitarias tienen en $\mathfrak{L}$ la longitud igual a $\sqrt{n}$;

b) las multiplicaciones por matrices conjugadas transpuestas X y $\overline{X}'$ originan en $\mathfrak{L}$ aplicaciones conjugadas;

c) la multiplicación de las matrices de $\mathfrak{L}$ por una matriz unitaria origina en $\mathfrak{L}$ una aplicación unitaria;

d) la multiplicación por una matriz hermitiana origina en $\mathfrak{L}$ una aplicación simétrica, mientras que la multiplicación por una matriz hermitiana antisimétrica origina una aplicación antisimétrica.

6. Demuéstrese que la suma de dos aplicaciones proyectivas $\mathcal{P}_{\mathfrak{A}}$ y $\mathcal{P}_{\mathfrak{B}}$ es una aplicación proyectiva cuando, y sólo cuando, $\mathcal{P}_{\mathfrak{A}}\mathcal{P}_{\mathfrak{B}} = \mathcal{O}$. Además, en este caso $\mathcal{P}_{\mathfrak{A}} + \mathcal{P}_{\mathfrak{B}} = \mathcal{P}_{\mathfrak{A}+\mathfrak{B}}$.

7. El producto de dos aplicaciones proyectivas $\mathcal{P}_{\mathfrak{A}}$ y $\mathcal{P}_{\mathfrak{B}}$ es una aplicación proyectiva cuando, y sólo cuando, $\mathcal{P}_{\mathfrak{A}}$ y $\mathcal{P}_{\mathfrak{B}}$ conmutan. Además, en este caso $\mathcal{P}_{\mathfrak{A}}\mathcal{P}_{\mathfrak{B}} = \mathcal{P}_{\mathfrak{A}\cap\mathfrak{B}}$.

8. Un subespacio $\mathfrak{A}$ está contenido en otro $\mathfrak{B}$ cuando, y sólo cuando, $\mathcal{P}_{\mathfrak{A}}\mathcal{P}_{\mathfrak{B}} = \mathcal{P}_{\mathfrak{A}}$.

9. Demuéstrese la fórmula $\mathcal{P}_{\mathfrak{A}^\perp} = \mathcal{E} - \mathcal{P}_{\mathfrak{A}}$.

10. Un subespacio $\mathfrak{A}$ es invariante respecto a una aplicación arbitraria $\mathcal{A}$ cuando, y sólo cuando, la correspondiente aplicación proyectiva $\mathcal{P} = \mathcal{P}_{\mathfrak{A}}$ verifica la relación $\mathcal{P}\mathcal{A}\mathcal{P} = \mathcal{P}\mathcal{A}$.

11. Si las aplicaciones $\mathcal{A}$ y $\mathcal{B}$ con las descomposiciones espectrales $\mathcal{A} = \sum \alpha_j \mathcal{P}_j$ y $\mathcal{B} = \sum \beta_k Q_k$ conmutan, toda aplicación $\mathcal{P}_j$ es conmutable con cualquier aplicación Q_k.

12. Si una aplicación normal $\mathcal{A}$ conmuta con cualquier aplicación lineal $\mathcal{B}$, la aplicación $\mathcal{A}$ conmuta también con $\mathcal{B}^*$.

13. Toda aplicación unitaria $\mathcal{U}$ de un espacio unitario complejo puede ser representada en la forma $\mathcal{U} = e^{i\mathcal{A}}$, donde $\mathcal{A}$ es una aplicación simétrica. Recíprocamente, la aplicación $e^{i\mathcal{A}}$ es unitaria cualquiera que sea la aplicación simétrica $\mathcal{A}$.

Capítulo VI

Formas bilineales y cuadráticas

Se supone que todos los espacios que aparecen en este capítulo son espacios sobre un cuerpo conmutativo (pero no sobre un cuerpo cualquiera).

§ 21. Formas bilineales

21.1. Transformación de formas. Un polinomio $F(\xi)$ en las variables $\xi_1, \ldots, \xi_n$ con coeficientes de un cuerpo conmutativo K se llama *forma de grado p-ésimo sobre K en* $\xi_1, \ldots, \xi_n$, si todos los términos de F son de un mismo grado p respecto al conjunto de las variables. Se llaman *lineales* las formas de primer grado, *cuadráticas* las de segundo grado, *cúbicas* las de tercer grado, etc.

Los problemas principales de la teoría de las formas son el problema del estudio de las leyes de variación de los coeficientes de las formas en las transformaciones lineales de las variables y el problema de la búsqueda de los tipos elementales a los que pueden ser reducidas las formas mediante estas transformaciones.

A veces, en lugar de una forma se considera un par de formas en las mismas variables y se plantea el problema de determinar una transformacion de las variables en la que ambas formas tomen la forma más sencilla posible. Este es el problema de un par de formas. Se pueden plantear problemas sobre ternas de formas, etc.

Al final de este capítulo daremos la interpretación geométrica del problema de transformación de formas, mientras que primero consideraremos el problema desde el punto de vista algebraico y daremos su solución para el caso de formas cuadráticas.

Escribiremos las fórmulas que relacionan las variables $\xi_1, \ldots, \xi_n$ con las variables nuevas $\xi_1', \ldots, \xi_n'$ en la forma

$$\xi_j = \xi_1'\tau_{1j} + \xi_2'\tau_{2j} + \ldots + \xi_n'\tau_{nj} \quad (j = 1, 2, \ldots, n) \qquad (1)$$

de acuerdo con el punto 5.1.

Supondremos siempre que *la matriz* $T = \|\tau_{ij}\|$ *es invertible* así que las relaciones (1) permiten siempre expresar las variables nuevas en términos de las antiguas. La matriz T se llama *matriz de transformación* de las variables ξ_i en las variables ξ'_j.

En general, las formas dadas suelen transformarse paulatinamente: primero se introducen unas variables nuevas mediante las fórmulas (1), después mediante fórmulas análogas se introducen en lugar de ξ'_j unas variables ξ''_j, etc. Sea $T = T_1$ la matriz de la transformación de las variables ξ en las variables ξ', sea T_2 la matriz de la transformación de ξ' en ξ'', etc. Introduciendo en las fórmulas (1) en lugar de ξ'_j sus expresiones en términos de ξ''_j, expresaremos las variables ξ_j linealmente en términos de ξ''_j. Los cálculos directos muestran que la matriz de la transformación de ξ en ξ'' es T_2T_1 (compárese con el p. 5.1). Aplicando este resultado varias veces, llegamos a la conclusión siguiente: *si a la transformación de las variables* ξ *en las variables* ξ' *le corresponde la matriz* T_1, *si a la transformación de* ξ' *en* ξ'' *le corresponde la matriz* T_2, etc., *entonces a la transformación resultante de las variables* ξ *en las variables* $\xi^{(m)}$ *le corresponde la matriz igual al producto* $T_mT_{m-1} \dots T_2T_1$ *de las matrices de las transformaciones intermedias.*

Supongamos, por ejemplo, que debemos reducir a la forma elemental la forma cuadrática

$$F = \xi_1^2 + \xi_3^2 - 2\xi_1\xi_2 - 2\xi_1\xi_3 + 10\xi_2\xi_3.$$

Tenemos

$$F = (\xi_1 - \xi_2 - \xi_3)^2 - \xi_2^2 + 8\xi_2\xi_3.$$

Introduciendo las variables nuevas

$$\xi'_1 = \xi_1 - \xi_2 - \xi_3, \quad \xi'_2 = \xi_2 \quad \text{y} \quad \xi'_3 = \xi_3,$$

obtenemos

$$F' = \xi_1'^2 - \xi_2'^2 + 8\xi'_2\xi'_3 = \xi_1'^2 - (\xi'_2 - 4\xi'_3)^2 + 16\xi_3'^2.$$

La transformación

$$\xi''_1 = \xi'_1, \quad \xi''_2 = \xi'_2 - 4\xi'_3 \quad \text{y} \quad \xi''_3 = 4\xi'_3$$

reduce la forma F a la forma elemental

$$F'' = \xi_1''^2 - \xi_2''^2 + \xi_3''^2.$$

Según la regla expuesta, la matriz de la transformación de ξ'' en ξ es

$$\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -4 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & -4 & 4 \end{bmatrix}.$$

Como segundo ejemplo consideremos el problema general de transformación de un sistema arbitrario de m formas lineales

$$\left.\begin{aligned} f_1 &= \xi_1\alpha_{11} + \xi_2\alpha_{21} + \ldots + \xi_n\alpha_{n1}, \\ &\cdots\cdots\cdots\cdots\cdots \\ f_m &= \xi_1\alpha_{1m} + \xi_2\alpha_{2m} + \ldots + \xi_n\alpha_{nm} \end{aligned}\right\} \qquad (2)$$

en n variables. La matriz $A = \|\alpha_{ij}\|$, que para $m \neq n$ resulta rectangular, se llama matriz de este sistema. Introduciendo en las formas dadas en lugar de las variables ξ_j sus expresiones en términos de ξ'_i definidas por (1), obtenemos un sistema de m formas lineales en las variables nuevas ξ'_i. El cálculo directo muestra que la matriz A_1 del sistema nuevo está ligada a la matriz A del sistema antiguo mediante la relación

$$A_1 = TA,$$

es decir, *al pasar a las variables nuevas la matriz de un sistema de formas lineales se multiplica a la izquierda por la matriz de la transformación.*

Para reducir el sistema (2) a la forma elemental escojamos entre las formas $f_1, \ldots, f_m, \xi_1, \ldots, \xi_n$ las n primeras linealmente independientes. Sean éstas las formas $f_{i_1}, \ldots, f_{i_r}, \xi_{i_{r+1}}, \ldots, \xi_{i_n}$. Es evidente que el número r es igual al rango de la matriz A. Podemos introducir en lugar de $\xi_1, \ldots, \xi_n$ unas variables ξ'_i tomando

$$\xi'_1 = f_{i_1}(\xi), \ \ldots, \ \xi'_r = f_{i_r}(\xi), \ \xi'_{r+1} = \xi_{r+1}, \ \ldots, \ \xi'_n = \xi_{i_n}$$

y después de ello el sistema dado (2) tomará la forma elemental requerida

$$\xi'_1, \ \ldots, \ \xi'_r, \ f'_{r+1}, \ \ldots, \ f'_m,$$

donde $f'_{r+1}, \ \ldots, \ f'_m$ son unas formas lineales en $\xi'_1, \ \ldots, \ \xi'_n$. En particular, si todas las formas iniciales eran linealmente independientes, la transformación de variables señalada las reducirá a la forma canónica $f'_1 = \xi'_1, \ \ldots, \ f'_m = \xi'_m$.

21.2. Equivalencia de formas bilineales. Frecuentemente en lugar de los polinomios en un sistema de variables $\xi_1, \ \ldots, \ \xi_n$ se consideran los polinomios en dos sistemas de variables, por ejemplo, $\xi_1, \ \ldots, \ \xi_n$ y $\eta_1, \ \ldots, \ \eta_s$, así como los polinomios en varios sistemas de variables. Un polinomio en varios sistemas de variables se llama *forma* si es homogéneo respecto a cada uno de los sistemas de variables por separado. Son de un interés especial las formas lineales respecto a cada uno de los sistemas de variables. Estas formas se llaman *bilineales*, si hay dos sistemas de variables, *trilineales*, si hay tres sistemas, y *polineales* en el caso general.

El número de variables de cada uno de los sistemas puede ser distinto. El problema de transformación de formas de varios siste-

mas de variables puede plantearse en diferentes aspectos, ya que se puede someter a una transformación lineal cada uno de los sistemas de variables independientemente de las transformaciones de los restantes sistemas y se puede también realizar con cada uno de estos sistemas transformaciones que estén ligadas entre sí de algún modo.

Se llaman *equivalentes* las formas que pueden ser reducidas una a otra mediante una selección independiente de transformaciones lineales de las variables. En cambio, se dice que las formas son *congruentes*, si todos los sistemas de variables contienen un mismo número de variables y las formas de estos sistemas pueden ser reducidas una a otra mediante transformaciones lineales—de una misma matriz—de cada uno de los sistemas. Está claro que las formas congruentes son siempre equivalentes. La recíproca, por supuesto, no tiene lugar en el caso general. Es fácil dar ejemplos de formas equivalentes que no son congruentes limitándose incluso al caso de formas bilineales. En este punto consideraremos el problema elemental de equivalencia de formas bilineales y en el punto siguiente examinaremos el problema sobre la congruencia de formas bilineales simétricas.

Una forma bilineal en dos sistemas de variables $\xi_1, \ldots, \xi_n$ y $\eta_1, \ldots, \eta_n$ es de la forma

$$F = \sum \alpha_{ij} \xi_i \eta_j \quad (i, j = 1, 2, \ldots, n).$$

La matriz $A = \|\alpha_{ij}\|$ formada por los coeficientes de la forma se llama *matriz de la forma* y el rango de la matriz A se llama *rango de la forma*.

Introduciendo las matrices de una fila

$$X = [\xi_1, \ldots, \xi_n] \quad \text{e} \quad Y = [\eta_1, \ldots, \eta_n],$$

podemos representar la forma F de modo siguiente

$$F = XAY'. \tag{3}$$

Supongamos ahora que debemos pasar de las variables ξ y η a unas variables nuevas ξ' y η' ligadas a las antiguas mediante las fórmulas

$$\xi_j = \sum \xi'_k \tau_{kj} \quad \text{y} \quad \eta_j = \sum \eta'_k \sigma_{kj}$$

o empleando la notación matricial

$$X = X_1 T \quad \text{e} \quad Y = Y_1 S, \tag{4}$$

donde $T = \|\tau_{ij}\|$, $S = \|\sigma_{ij}\|$ y

$$X_1 = [\xi'_1, \ldots, \xi'_n] \quad \text{e} \quad Y_1 = [\eta'_1, \ldots, \eta'_n].$$

Introduciendo en (3) las expresiones (4) para X e Y, obtenemos

$$F = X_1 TAS'Y'_1 = X_1 A_1 Y'_1,$$

donde A_1 es la matriz de la forma transformada.

Por consiguiente, si en una forma bilineal de matriz A se realiza una transformación de matriz T del primer sistema de variables y una transformación de matriz S del segundo sistema de variables, se obtiene una forma bilineal de matriz

$$A_1 = TAS'. \tag{5}$$

Hemos explicado ya que las formas bilineales que se obtienen una de otra mediante transformaciones lineales de las variables se llaman equivalentes. Por otra parte, según el p.13.4 unas matrices A y A_1 se llaman equivalentes sobre un cuerpo conmutativo K, si existen unas matrices regulares P y Q formadas por elementos de K tales que $A_1 = PAQ$. Comparando con la fórmula (5), vemos que *para la equivalencia de unas formas bilineales sobre un cuerpo conmutativo arbitrario K es necesario y suficiente que sus matrices sean equivalentes.*

Según el p. 13.4, todas las matrices cuadradas de un orden dado n y de un rango dado r son equivalentes entre sí sobre el cuerpo conmutativo K y son equivalentes a una matriz de tipo $E_r \dotplus O_{n-r}$, donde E_r es la matriz unidad de orden r y O_{n-r} es la matriz nula de orden $n-r$. Aplicando esto a las formas bilineales obtenemos el siguiente resultado que resuelve totalmente el problema de equivalencia de las formas bilineales:

Para la equivalencia de unas formas bilineales sobre un cuerpo conmutativo arbitrario es necesario y suficiente que coincidan los órdenes y los rangos respectivos de las matrices de estas formas.

Las formas de determinante diferente de cero se llaman *regulares* y las demás se llaman *singulares*. El resultado obtenido acerca de la equivalencia de formas significa que todas las formas bilineales regulares en sistemas de n variables son equivalentes a la forma

$$\xi_1\eta_1 + \xi_2\eta_2 + \ldots + \xi_n\eta_n,$$

mientras que todas las formas singulares son equivalentes a formas de tipo

$$\xi_1\eta_1 + \xi_2\eta_2 + \ldots + \xi_r\eta_r \qquad (r = 0, 1, 2, \ldots, n-1),$$

donde r es el rango de la forma.

Si el campo principal es el cuerpo de los números complejos, suelen considerarse, además de las formas bilineales corrientes, *las formas de Hermite*, es decir, las formas de tipo

$$F = \sum \alpha_{ij}\xi_i\overline{\eta}_j,$$

donde la raya superior significa que se pasa a los valores conjugados. La notación matricial de una forma bilineal hermitiana de matriz $A = \|\alpha_{ij}\|$ es

$$F = XA\overline{Y}',$$

y la matriz A_1 de la forma hermitiana nueva que resulta de F al realizar la transformación (4) de las variables es igual a

$$A_1 = TA\overline{S}'. \tag{6}$$

De aquí obtenemos, al igual que antes, que *todas las formas bilineales hermitianas en sistemas de n variables son equivalentes a la forma*

$$\xi_1\overline{\eta}_1 + \xi_2\overline{\eta}_2 + \ldots + \xi_r\overline{\eta}_r,$$

donde r es el rango de la forma dada.

21.3. Congruencia de formas bilineales simétricas. Hemos señalado anteriormente que las formas bilineales en dos sistemas compuestos por un mismo número de variables se llaman congruentes si se obtienen una de otra mediante transformaciones lineales de matrices idénticas de ambos sistemas.

Tomando $S = T$ en la fórmula (5), llegamos a la conclusión de que *al someter ambos sistemas de variables de una forma bilineal de matriz A a una transformación lineal de matriz T, la matriz de la forma nueva será*

$$A_1 = TAT'. \tag{7}$$

Unas matrices A y A_1 se llaman *congruentes*, si están ligadas por una relación de tipo (7), donde T es una matriz regular adecuada. Por consiguiente, las formas bilineales son congruentes si, y sólo si, son congruentes sus matrices.

Una forma bilineal cuya matriz es simétrica o antisimétrica también se llama simétrica o antisimétrica, respectivamente.

Si una forma bilineal dada es simétrica o antisimétrica, la misma propiedad tendrán todas las formas congruentes.

Efectivamente si $A' = \pm A$, de (7) resulta

$$A_1' = \pm TA'T' = \pm A_1.$$

Considerando análogamente las formas bilineales hermitianes, obtenemos de (6) que la matriz A_1 de la forma hermitiana nueva que resulta al aplicar a la forma de matriz A una transformación lineal de matriz T de ambos sistemas de variables es igual a

$$A_1 = TA\overline{T}'. \tag{8}$$

Unas matrices A y A_1 se llaman *hermitianas congruentes*, si están ligadas por la relación (8) mediante una matriz regular T. Por lo tanto, la congruencia de las formas hermitianas equivale a la congruencia de hermitiana sus matrices.

Una forma hermitiana se llama *simétrica* si su matriz es hermitiana simétrica, es decir, si $\overline{A}' = A$. De (8) se desprende directamente que si una forma hermitiana dada es simétrica, todas las formas congruentes de la misma son también simétricas.

Puesto que la congruencia de matrices implica la equivalencia de las mismas, para la congruencia de unas formas es necesario que coincidan sus rangos. Sin embargo, esta condición no es suficiente. Las condiciones suficientes en el caso más general serán examinadas en el p. 23.3; ahora daremos estas condiciones para los casos más importantes solamente.

Sea, ante todo, K el cuerpo de los números reales y sea A una matriz simétrica. Existe entonces, de acuerdo con el p. 19.4, una matriz unitaria real U tal que la matriz $A_1 = UAU^{-1}$ tendrá la forma diagonal. Pero si U es unitaria, tenemos $U\overline{U}' = E$, de donde, debido a que las matrices son reales, se deduce que $U^{-1} = U'$, es decir, $A_1 = UAU'$. Por consiguiente, A es *congruente* de una matriz diagonal A_1. Luego, hemos demostrado el teorema siguiente:

TEOREMA 1. *Toda forma bilineal simétrica real puede ser reducida, mediante una adecuada transformación unitaria real de las variables, a la forma*

$$\alpha_1\xi_1\eta_1 + \alpha_2\xi_2\eta_2 + \ldots + \alpha_r\xi_r\eta_r, \tag{9}$$

donde r es el rango de la forma y $\alpha_1, \ldots, \alpha_r$ son los números característicos diferentes de cero de la matriz de la forma. En particular, para la equivalencia unitaria de unas formas bilineales simétricas reales es necesario y suficiente que coincidan los polinomios característicos de las matrices de las formas.

Este teorema de más de lo que pretendíamos obtener. Significa que la reducción a la forma diagonal se puede alcanzar mediante una transformación *unitaria* real de las variables. Si no es necesario que la transformación de las variables sea unitaria, podemos continuar la reducción a la forma elemental. Es decir, supongamos que hemos reducido ya la forma al tipo diagonal (9). Cambiemos ahora la numeración de las variables de manera que primero aparezcan los términos de coeficientes positivos y después los de coeficientes negativos. Supongamos, por ejemplo, que $\alpha_1, \ldots, \alpha_s$ son positivos y que $\alpha_{s+1}, \ldots, \alpha_r$ son negativos. Tomando entonces

$$\xi_j' = \sqrt{\alpha_j}\,\xi_j, \qquad \eta_j' = \sqrt{\alpha_j}\,\eta_j \qquad (j = 1, \ldots, s),$$
$$\xi_k' = \sqrt{-\alpha_k}\,\xi_k, \quad \eta_k' = \sqrt{-\alpha_k}\,\eta_k \qquad (k = s+1, \ldots, r),$$

podemos reducir la forma dada a la forma

$$\xi_1'\eta_1' + \ldots + \xi_s'\eta_s' - \xi_{s+1}'\eta_{s+1}' - \ldots - \xi_r'\eta_r'. \tag{10}$$

La diferencia

$$\sigma = s - (r - s) = 2s - r$$

se llama *signatura* de la forma (10). Es evidente que la forma (10) se determina totalmente por su rango y su signatura, ya que $s = \frac{1}{2}(\sigma + r)$. El hecho de que la signatura no depende de cómo se

reduce la forma dada a la forma (i0) y, por consiguiente, se determina unívocamente por la forma inicial, constituye el contenido de la así llamada *ley de inercia* que será examinada detalladamente en el p. 22.3.

Para las formas simétricas hermitianas la situación es totalmente análoga. Sea F una forma hermitiana de matriz hermitiana simétrica A. De acuerdo con el teorema 4a (p.19.4), existe una matriz unitaria U tal que la matriz $A_1 = UAU^{-1}$ será diagonal real. Como U es unitaria, tenemos $U^{-1} = \overline{U}'$, de donde $A_1 = UA\overline{U}'$. Hemos demostrado, pues, el siguiente teorema:

TEOREMA 2. *Toda forma bilineal hermitiana simétrica puede ser reducida, mediante una transformación unitaria de las variables, a la forma diagonal*

$$\alpha_1\xi_1\overline{\eta}_1 + \alpha_2\xi_2\overline{\eta}_2 + \ldots + \alpha_r\xi_r\overline{\eta}_r$$

con coeficientes reales. Los valores $\alpha_1, \ldots, \alpha_r$ *son las raíces diferentes de cero del polinomio característico de la matriz de la forma y por ello para la congruencia unitaria de las formas hermitianas simétricas es necesario y suficiente que coincidan los polinomios característicos de estas formas.*

Si en lugar de la congruencia unitaria se considera la congruencia respecto a transformaciones lineales arbitrarias, el proceso de reducción de una forma puede ser continuado como ha sido señalado anteriormente y así obtendremos una forma de tipo

$$\xi_1\overline{\eta}_1 + \ldots + \xi_s\overline{\eta}_s - \xi_{s+1}\overline{\eta}_{s+1} - \ldots - \xi_r\overline{\eta}_r. \qquad (11)$$

Toda forma hermitiana simétrica puede ser reducida, por consiguiente, a una de estas $n+1$ formas ($r = 0, 1, \ldots, n$). Otra vez la no congruencia de las formas (11) para diferentes valores de s se desprende de la ley de inercia mencionada anteriormente.

Ejemplos y problemas

1. Demuéstrese que el sistema de formas lineales en ξ_1, ξ_2, ξ_3 y ξ_4

$$\xi_1+\xi_2,\ \xi_2+\xi_3,\ \xi_3+\xi_4 \quad \text{y} \quad \xi_1+\xi_4$$

es equivalente al sistema

$$\xi_1+\xi_2+\xi_3,\ \xi_2+\xi_3+\xi_4,\ \xi_1+\xi_3+\xi_4 \quad \text{y} \quad 2\xi_1+\xi_3$$

y no es equivalente al sistema

$$\xi_1-\xi_2+\xi_3,\ \xi_2+\xi_3-\xi_4,\ \xi_1+\xi_3 \quad \text{y} \quad 2\xi_1+3\xi_2-\xi_4.$$

2. Demuéstrese que un sistema de formas lineales

$$f_i = \xi_1\alpha_{1i} + \xi_2\alpha_{2i} + \ldots + \xi_n\alpha_{ni} \qquad (i = 1, \ldots, m) \qquad (12)$$

es equivalente a otro sistema

$$g_i = \xi_1\beta_{1i} + \xi_2\beta_{2i} + \ldots + \xi_n\beta_{ni} \qquad (i = 1, \ldots, m) \qquad (13)$$

cuando, y sólo cuando, el sistema de vectores

$$a_i=[\alpha_{1i}, \alpha_{2i}, \ldots, \alpha_{ni}] \quad (i=1, \ldots, m) \tag{14}$$

del espacio lineal de los vectores filas puede ser convertido, mediante una adecuada transformación lineal regular de este espacio, en el sistema de vectores

$$b_i=[\beta_{1i}, \beta_{2i}, \ldots, \beta_{ni}] \quad (i=1, \ldots, m). \tag{15}$$

3. Demuéstrese que el sistema de formas lineales (12) puede ser reducido, mediante una transformación *unitaria* de las variables, en el sistema (13) cuando, y sólo cuando, el sistema de vectores (14) del *espacio unitario de filas* puede ser convertido, mediante una adecuada aplicación unitaria de este espacio, en el sistema de vectores (15).

4. Demuéstrese que para la equivalencia de formas bilineales en sistemas que contienen diferente número de variables es necesario y suficiente que coincidan las dimensiones (es decir, el número de filas y el número de columnas) y los rangos de las matrices de las formas. En particular, las formas bilineales en dos sistemas de variables $\xi_1, \ldots, \xi_n$ y $\eta_1, \ldots, \eta_m$ $(n > m)$ de rango r son equivalentes a la forma $\xi_1\eta_1+\ldots+\xi_r\eta_r$.

§ 22. Formas cuadráticas

22.1. Congruencia. Según la definición general, una forma cuadrática en las variables $\xi_1, \ldots, \xi_n$ es un polinomio homogéneo de segundo grado en estas variables. Toda forma cuadrática en las variables indicadas puede ser representada unívocamente en la siguiente forma simétrica

$$F(\xi)=\sum \alpha_{ij}\xi_i\xi_j \quad (\alpha_{ij}=\alpha_{ji}). \tag{1}$$

La matriz $A=\|\alpha_{ij}\|$ se llama matriz de la forma cuadrática y la forma bilineal simétrica

$$F(\xi, \eta)=\sum \alpha_{ij}\xi_i\eta_j$$

en dos sistemas de variables, que tiene la misma matriz que la forma cuadrática, se llama forma *polar* de esta última. Identificando en la forma polar el primero y el segundo sistemas de variables obtenemos la forma cuadrática inicial. Así se establece una correspondencia biyectiva entre las formas cuadráticas y las formas bilineales simétricas. Por ejemplo, si

$$F(\xi)=\xi_1^2-\xi_2^2+3\xi_1\xi_3-6\xi_2\xi_3,$$

la notación simétrica de $F(\xi)$ es

$$F(\xi)=\xi_1^2-\xi_2^2+\frac{3}{2}\xi_1\xi_3+\frac{3}{2}\xi_3\xi_1-3\xi_2\xi_3-3\xi_3\xi_2$$

y la forma polar correspondiente es

$$F(\xi, \eta)=\xi_1\eta_1-\xi_2\eta_2+\frac{3}{2}\xi_1\eta_3+\frac{3}{2}\xi_3\eta_1-3\xi_2\eta_3-3\xi_3\eta_2.$$

Se dice que una forma cuadrática es diagonal si su matriz es diagonal, es decir, si la forma contiene sólo términos con los cuadrados de las variables.

Realizando en la forma (1) una transformación lineal de las variables de matriz T, obtendremos una forma nueva (p. 21.1)

$$F(\xi) = XAX' = X_1TAT'X_1'$$
$$(X = [\xi_1, \ldots, \xi_n] \quad \text{y} \quad X_1 = [\xi_1', \ldots, \xi_n'])$$

de matriz

$$A_1 = TAT'.$$

Por consiguiente, la ley de variación de la matriz de una forma cuadrática es la misma que para la correspondiente forma bilineal polar. De aquí se deduce que unas formas cuadráticas son congruentes cuando, y sólo cuando, son congruentes las correspondientes formas polares y del teorema 1 del parágrafo anterior obtenemos directamente el teorema siguiente:

TEOREMA 1. *Todo forma cuadrática real puede ser reducida mediante una adecuada transformación ortogonal real de las variables, a la forma diagonal*

$$\alpha_1\xi_1^2 + \alpha_2\xi_2^2 + \ldots + \alpha_r\xi_r^2, \tag{2}$$

donde r es el rango de la forma inicial y $\alpha_1, \ldots, \alpha_r$ son los números característicos diferentes de cero de la matriz de la forma. En particular, para la congruencia ortogonal de unas formas cuadráticas reales es necesario y suficiente que coincidan los polinomios característicos de las formas.

Si se aceptan también transformaciones no ortogonales de las variables, es posible continuar la reducción: si $\alpha_1, \ldots, \alpha_s$ son positivos y $\alpha_{s+1}, \ldots, \alpha_r$ son negativos, la forma (2) se reduce mediante la sustitución

$$\xi_i' = \sqrt{\alpha_i}\,\xi_i, \quad \xi_j' = \sqrt{-\alpha_j}\,\xi_j \quad (i = 1, \ldots, s; \; j = s+1, \ldots, r) \tag{3}$$

a la forma

$$\xi_1'^2 + \ldots + \xi_s'^2 - \xi_{s+1}'^2 - \ldots - \xi_r'^2.$$

Por consiguiente, a esta forma puede ser reducida, sobre el cuerpo de los números reales, cualquier forma cuadrática. El caso de otros cuerpos conmutativos será considerado en el punto siguiente.

Una expresión de tipo

$$F(\xi) = \sum \alpha_{ij}\xi_i\bar{\xi}_j = XA\bar{X}' \quad (X = [\xi_1, \ldots, \xi_n]),$$

donde $\alpha_{ij} = \bar{\alpha}_{ji}$ $(i, j = 1, \ldots, n)$, se llama *forma cuadrática hermitiana* en las variables $\xi_1, \ldots, \xi_n$ de matriz $A = \|\alpha_{ij}\|$. La forma bilineal simétrica hermitiana $F(\xi, \eta) = \sum \alpha_{ij}\xi_i\bar{\eta}_j$ se llama *forma polar* de $F(\xi)$. Las leyes de variación de las matrices de una forma

cuadrática hermitiana y de su forma polar coinciden. Por esto la congruencia de las formas cuadráticas hermitianas equivale a la congruencia de sus formas polares y del teorema sobre la reducción de formas bilineales simétricas hermitianas (p. 21.3) se deduce el teorema siguiente:

TEOREMA 2. *Toda forma cuadrática hermitiana puede ser reducida, mediante una transformación unitaria de las variables, a la forma*

$$\alpha_1\xi_1\bar{\xi}_1 + \alpha_2\xi_2\bar{\xi}_2 + \ldots + \alpha_r\xi_r\bar{\xi}_r, \qquad (4)$$

donde r es el rango y $\alpha_1, \ldots, \alpha_r$ son los números característicos no nulos de la matriz de la forma dada. Para la congruencia unitaria de unas formas cuadráticas hermitianas es necesario y suficiente que coincidan los polinomios característicos de las matrices de las formas.

Mediante una transformación ulterior de las variables de tipo (3), que ya no será unitaria, la forma (4) puede ser reducida a la forma

$$\xi_1\bar{\xi}_1 + \ldots + \xi_s\bar{\xi}_s - \xi_{s+1}\bar{\xi}_{s+1} - \ldots - \xi_r\bar{\xi}_r$$

que es en este caso la elemental.

22.2. Algoritmo de Lagrange. Uno de los métodos más simples de reducción de una forma cuadrática a la forma diagonal es el así llamado método de Lagrange que será considerado aquí. Se puede aceptar que el campo principal es un cuerpo conmutativo cualquiera de característica diferente de 2.

Supongamos que la forma (1) debe ser reducida a la forma diagonal. Pueden darse dos casos: a) la forma contiene el cuadrado de al menos una variable y b) la forma no contiene los cuadrados de las variables.

a) Supongamos, por ejemplo, que $\alpha_{11} \neq 0$. Representando la forma del modo siguiente

$$F = \alpha_{11}\xi_1^2 + 2\alpha_{12}\xi_1\xi_2 + \ldots + 2\alpha_{1n}\xi_1\xi_n + \sum_{\lambda>1,\ \mu>1} \alpha_{\lambda\mu}\xi_\lambda\xi_\mu =$$
$$= \alpha_{11}^{-1}(\alpha_{11}\xi_1 + \alpha_{12}\xi_2 + \ldots + \alpha_{1n}\xi_n)^2 - \alpha_{11}^{-1}(\alpha_{12}^2\xi_2^2 + 2\alpha_{12}\alpha_{13}\xi_2\xi_3 + \ldots$$
$$\ldots + \alpha_{1n}^2\xi_n^2) + \sum_{\lambda>1,\ \mu>1} \alpha_{\lambda\mu}\xi_\lambda\xi_\mu = \alpha_{11}^{-1}(\alpha_{11}\xi_1 + \alpha_{12}\xi_2 + \ldots$$
$$\ldots + \alpha_{1n}\xi_n)^2 + F_1(\xi_2, \ldots, \xi_n),$$

donde F_1 es una forma cuadrática en $\xi_2, \ldots, \xi_n$, y realizando la sustitución

$$\xi_1' = \alpha_{11}\xi_1 + \alpha_{12}\xi_2 + \ldots + \alpha_{1n}\xi_n,$$
$$\xi_i' = \xi_i \quad (i = 2, 3, \ldots, n),$$

podemos reducir F a la forma

$$F = \alpha_{11}^{-1}\xi_1'^2 + F_1,$$

donde F_1 no depende de ξ_1'.

b) Supongamos que $\alpha_{11} = \ldots = \alpha_{nn} = 0$ y que, por ejemplo, $\alpha_{12} \neq 0$. Representando la forma F del modo siguiente

$$F = 2\xi_1 (\alpha_{12}\xi_2 + \ldots + \alpha_{1n}\xi_n) + F_1 (\xi_2, \ldots, \xi_n)$$

y realizando la transformación

$$\begin{aligned} \xi_2' &= \alpha_{12}\xi_2 + \ldots + \alpha_{1n}\xi_n - \xi_1 \\ \xi_1' &= \xi_i \quad (i = 1, 3, 4, \ldots, n), \end{aligned}$$

obtenemos la forma

$$F = 2\xi_1' (\xi_1' + \xi_2') + F_1 = 2\xi_1'^2 + 2\xi_1'\xi_2' + 2\alpha_{12}^{-1}\alpha_{23}\xi_1'\xi_3' + \ldots$$

que contiene el cuadrado de la variable ξ_1'.

Por consiguiente, aplicando el proceso a) y complementándolo, en los casos necesarios, con el proceso b) podemos reducir la forma dada a la forma diagonal.

Ejemplo. Es necesario reducir a la forma diagonal la forma

$$F = \xi_1^2 + 4\xi_2^2 + 8\xi_3^2 - \xi_4^2 - 4\xi_1\xi_2 + 6\xi_1\xi_3 - 12\xi_2\xi_3 + 2\xi_2\xi_4 + \xi_3\xi_4 - \xi_4\xi_5.$$

Tenemos

$$F = (\xi_1 - 2\xi_2 + 3\xi_3)^2 - \xi_2^2 - \xi_3^2 + 2\xi_2\xi_4 + \xi_3\xi_5 - \xi_4\xi_5.$$

Realizando la transformación

$$\xi_1' = \xi_1 - 2\xi_2 + 3\xi_3, \quad \xi_i' = \xi_i \quad (i > 1),$$

obtenemos la forma

$$F_1 = \xi_1'^2 - \xi_2'^2 + 2\xi_2'\xi_4' - \xi_4'^2 + \xi_3'\xi_5' - \xi_4'\xi_5' = \xi_1'^2 - (\xi_2' - \xi_4')^2 + \xi_3'\xi_5' - \xi_4'\xi_5',$$

que mediante la transformación

$$\eta_3 = \xi_2' - \xi_4', \quad \eta_i' = \xi_i' \quad (i \neq 3)$$

se reduce a la forma

$$F_2 = \eta_1'^2 - \eta_2'^2 + \eta_3'\eta_5' - \eta_4'\eta_5'.$$

Ahora, de acuerdo con b), realizamos la transformación

$$\eta_4' = \eta_3' - \eta_4' - \eta_5', \quad \eta_i' = \eta_i' \quad (i \neq 4),$$

obteniendo así la forma

$$F_3 = \eta_1'^2 - \eta_3'^2 + (\eta_4' + \eta_5')\eta_5' = \eta_1'^2 - \eta_3'^2 + \left(\eta_5' + \frac{1}{2}\eta_4'\right)^2 - \frac{1}{4}\eta_4'^2.$$

Esta forma mediante la transformación

$$\zeta_2 = \eta_2' + \frac{1}{2}\eta_4', \quad \zeta_i = \eta_i' \quad (i = 1, 3, 4), \quad \zeta_5 = \eta_5'$$

se reduce a la forma diagonal

$$F_4 = \zeta_1^2 + \zeta_3^2 - \zeta_5^2 - \frac{1}{4}\zeta_4^2.$$

Para hallar la matriz T de la transformación de las variables ξ_i en las variables iniciales ξ_i es suficiente, por lo visto en el p. 21.1, multiplicar las matrices de las transformaciones intermedias.

Si en vez de una forma cudrática es necesario reducir a la forma diagonal una forma *bilineal* simétrica, sustituimos la forma bilineal por la cuadrática correspondiente, reducimos esta última a la forma diagonal y determinamos la matriz de la transformación. Debido a la relación que existe entre las formas cuadráticas y bilineales, esta transformación también reducirá a la forma diagonal la forma bilineal dada.

Para concluir, consideremos el problema sobre la reducción a la forma elemental de una forma bilineal antisimétrica

$$F = \sum \alpha_{ij} \xi_i \eta_j \quad (\alpha_{ij} = -\alpha_{ji})$$

cuyos coeficientes pertenecen a un cuerpo conmutativo cualquiera.

Si todos los coeficientes son iguales a cero, la forma está ya reducida. En el caso contrario aceptaremos, por ejemplo, que $\alpha_{12} \neq 0$. Escribiendo la forma del modo siguiente

$$F = \xi_1 (\alpha_{12}\eta_2 + \ldots + \alpha_{1n}\eta_n) - \eta_1 (\alpha_{12}\xi_2 + \ldots + \alpha_{1n}\xi_n) + \\ + R(\xi_2, \ldots, \xi_n, \eta_2, \ldots, \eta_n)$$

y realizando la sustitución

$$\xi_2' = \alpha_{12}\xi_2 + \ldots + \alpha_{1n}\xi_n, \quad \eta_2' = \alpha_{12}\eta_2 + \ldots + \alpha_{1n}\eta_n,$$
$$\xi_i' = \xi_i, \quad \eta_i' = \eta_i \quad (i = 1, 3, \ldots, n),$$

obtenemos la forma

$$F_1 = \xi_1'\eta_2' - \xi_2'\eta_1' + R_1(\xi_2', \ldots, \xi_n', \eta_2', \ldots, \eta_n').$$

Ahora pueden darse dos casos: a) el resto R_1 no contiene ξ_2' y, por consiguiente, tampoco contiene η_2' y b) el resto contiene ξ_2'. En el caso a) aplicamos el proceso solamente al resto, ya que éste no depende del término que hemos despejado $\xi_1'\eta_2' - \xi_2'\eta_1'$. En el caso b) representamos la forma del modo siguiente

$$F_1 = \xi_1'\eta_2' - \xi_2'\eta_1' + \xi_3'(\alpha_{23}'\eta_2' + \ldots + \alpha_{2n}'\eta_n') - \\ - \eta_2'(\alpha_{23}'\xi_3' + \ldots + \alpha_{2n}'\xi_n') + R_2(\xi_3', \ldots) = \\ = (\xi_1' - \alpha_{23}'\xi_3' - \ldots - \alpha_{2n}'\xi_n')\eta_2' - \\ - (\eta_1' - \alpha_{23}'\eta_3' - \ldots - \alpha_{2n}'\eta_n')\xi_2' + R_2(\xi_3', \ldots)$$

y realizamos la sustitución

$$\xi_1'' = \xi_1' - \alpha_{23}'\xi_3' - \ldots - \alpha_{2n}'\xi_n',$$
$$\eta_1'' = \eta_1' - \alpha_{23}'\eta_3' - \ldots - \alpha_{2n}'\eta_n',$$
$$\xi_i'' = \xi_i', \quad \eta_i'' = \eta_i' \quad (i = 2, \ldots, n),$$

obteniendo como resultado la forma descompuesta

$$\xi_1''\eta_2'' - \xi_2''\eta_1'' + R_2(\xi_3'', \ldots, \eta_3'', \ldots).$$

Aplicando el proceso expuesto al resto, después al resto nuevo, etc., reduciremos la forma a

$$\xi_1\eta_2 - \xi_2\eta_1 + \xi_3\eta_4 - \xi_4\eta_3 + \ldots + \xi_{2r-1}\eta_{2r} - \xi_{2r}\eta_{2r-1},$$

donde $2r$ es el rango de la forma reducida y, por consiguiente, también de la inicial. En particular, obtenemos de nuevo que *el rango de una matriz antisimétrica es siempre un número par.*

Ejemplo. Redúzcase a la forma elemental la forma

$$F = \xi_1\eta_2 - \xi_2\eta_1 + \xi_1\eta_3 - \xi_3\eta_1 - 2\xi_1\eta_4 + 2\xi_4\eta_1 + 3\xi_2\eta_3 - 3\xi_3\eta_2 + \\ + \xi_2\eta_4 - \xi_4\eta_2 - 4\xi_3\eta_4 + 4\xi_4\eta_3.$$

De acuerdo con la regla, realizamos la primera transformación

$$\xi_2' = \xi_2 + \xi_3 - 2\xi_4, \quad \eta_4' = \eta_2 + \eta_3 - 2\eta_4,$$
$$\xi_i' = \xi_i, \quad \eta_i' = \eta_i \quad (i = 1, 3, 4),$$

después de la cual la forma se convierte en

$$F_1 = \xi_1'\eta_2' - \xi_2'\eta_1' + 3\xi_2'\eta_3' - 3\xi_3'\eta_2' + \xi_2'\eta_4' - \xi_4'\eta_2' - 4\xi_3'\eta_4' + 4\xi_4'\eta_3'.$$

Realizando ahora la transformación

$$\xi_1'' = \xi_1' - 3\xi_3' - \xi_4', \quad \eta_1'' = \eta_1' - 3\eta_3' - \eta_4',$$
$$\xi_i'' = \xi_i', \quad \eta_i'' = \eta_i' \quad (i = 2, 3, 4),$$

obtenemos la forma elemental

$$F_2 = \xi_1''\eta_2'' - \xi_2''\eta_1'' - 4\xi_3''\eta_4'' + 4\xi_4''\eta_3''.$$

Desde el punto de vista práctico el problema sobre la reducción de las formas cuadráticas a la forma elemental se descompone en dos momentos: la determinación de la forma elemental definitiva y la determinación de la matriz de la transformación de las variables necesaria para reducir la forma a la forma elemental. Si la reducción se realiza mediante transformaciones lineales arbitrarias, ambos problemas quedan resueltos al aplicar el algoritmo de Lagrange.

La situación es más compleja si las formas se reducen mediante transformaciones *unitarias* de las variables. Sea dada una forma hermitiana de matriz A. En el espacio unitario auxiliar de vectores filas la multiplicación de las filas por la matriz A representa una aplicación lineal simétrica $\mathcal{A}$ de matriz A en la base elemental (p. 4.1). Debemos determinar una base ortonormal en la que la matriz A de la aplicación sea de forma diagonal. Para ello hallamos los números característicos $\alpha_1, \ldots, \alpha_n$ de la matriz A que se obtienen resolviendo la ecuación característica de A. Después,

resolviendo el sistema de ecuaciones lineales

$$[\xi_1, \ldots, \xi_n]\, A = \alpha_i\, [\xi_1, \ldots, \xi_n]$$

respecto a las incógnitas $\xi_1, \ldots, \xi_n$, encontramos n vectores propios linealmente independientes

$$x_i = [\xi_1^{(i)}, \ldots, \xi_n^{(i)}] \quad (i = 1, \ldots, n)$$

de la aplicación $\mathcal{A}$. Normalizando estos vectores, obtenemos una base ortonormal en la que la matriz A y, por consiguiente, la forma hermitiana inicial adquieren la forma diagonal con los elementos diagonales $\alpha_1, \ldots, \alpha_n$. Si los vectores x_i son ya normalizados, la matriz requerida de la transformación será $T = \| \xi_j^{(i)} \|$.

22.3. Ley de inercia de formas cuadráticas. En los resultados del p. 22.1. acerca de la reducción de formas cuadráticas reales, así como de formas hermitianas complejas, a la forma diagonal hemos dejado una laguna: no hemos aclarado si pueden ser congruentes formas de diferente signatura. El teorema que sigue llena esta laguna.

TEOREMA 3 (ley de inercia). *Toda forma cuadrática real puede ser reducida a la forma diagonal* $\alpha_1 \xi_1^2 + \ldots + \alpha_r \xi_r^2$ *mediante un número infinito de transformaciones de las variables. Sin embargo, aun cuando los propios coeficientes* $\alpha_1, \ldots, \alpha_r$ *pueden depender de la transformación que se emplee, el número de los coeficientes positivos y el número de los coeficientes negativos que aquí figuran no dependen de las transformaciones señaladas y, por consiguiente, se determinan unívocamente por la propia forma inicial.*

Supongamos, al contrario, que una forma cuadrática real

$$\alpha_1 \xi_1^2 + \ldots + \alpha_s \xi_s^2 - \alpha_{s+1} \xi_{s+1}^2 - \ldots - \alpha_r \xi_r^2 \quad (\alpha_i > 0)$$

se convierte, mediante la transformación de las variables $\xi_i = \sum \xi'_\lambda \tau_{\lambda i}$, de nuevo en la forma diagonal

$$\beta_1 \xi_1'^2 + \ldots + \beta_t \xi_t'^2 - \beta_{t+1} \xi_{t+1}'^2 - \ldots - \beta_r \xi_r'^2 \quad (\beta_i > 0),$$

pero siendo $s < t$. Esto significa que la igualdad

$$\alpha_1 \xi_1^2 + \ldots + \alpha_s \xi_s^2 - \alpha_{s+1} \xi_{s+1}^2 - \ldots - \alpha_r \xi_r^2 =$$
$$= \beta_1 \xi_1'^2 + \ldots + \beta_t \xi_t'^2 - \beta_{t+1} \xi_{t+1}'^2 - \ldots - \beta_r \xi_r'^2$$

se convierte en una identidad al sustituir las variables ξ_i por sus expresiones en términos de las variables ξ'_i. Representemos esta identidad en la forma

$$\alpha_1 \xi_1^2 + \ldots + \alpha_s \xi_s^2 + \beta_{t+1} \xi_{t+1}'^2 + \ldots + \beta_r \xi_r'^2 =$$
$$= \beta_1 \xi_1'^2 + \ldots + \beta_t \xi_t'^2 + \alpha_{s+1} \xi_{s+1}^2 + \ldots + \alpha_r \xi_r^2 \qquad (5)$$

y consideremos el sistema de ecuaciones

$$\xi_1 = 0, \ldots, \xi_s = 0, \xi'_{t+1} = 0, \ldots, \xi'_n = 0, \qquad (6)$$

donde por $\xi_1, \ldots, \xi_s$ se comprenden sus expresiones en términos de $\xi'_1, \ldots, \xi'_n$. El sistema (6) es un sistema de ecuaciones lineales homogéneas respecto a las incógnitas $\xi'_1, \ldots, \xi'_n$, siendo el número de ecuaciones $s+(n-t)=n-(t-s)$ indudablemente menor que el número de incógnitas, ya que $t>s$. Pero en estas condiciones el sistema (6) debe tener al menos una solución no nula

$$\xi'_1=\gamma_1, \ \ldots, \ \xi'_t=\gamma_t, \ \xi'_{t+1}=0, \ \ldots, \ \xi'_n=0. \tag{7}$$

Introduciéndola en la identidad (5), obtenemos

$$\beta_1\gamma_1^2+\ldots+\beta_t\gamma_t^2+\alpha_{s+1}\xi_{s+1}^2+\ldots+\alpha_r\xi_r^2=0. \tag{8}$$

Los números β_i y α_j son positivos y los números γ_i^2 y ξ_j^2 son no negativos y por ello de (8) resulta que $\gamma_1=\ldots=\gamma_t=0$ lo que contradice a que la solución (7) es no trivial.

La ley de inercia tiene lugar también, con el mismo enunciado, en el caso de formas cuadráticas hermitianas. La demostración no difiere de la que acabamos de realizar.

22.4. Formas de signo constante. Una forma cuadrática real $F(\xi)=\sum \alpha_{ij}\xi_i\xi_j$ se llama *no negativa* si su valor es no negativo cualesquiera que sean los valores reales de las variables ξ_i. La forma se llama *definida positiva* si su valor es estrictamente positivo para cualquier sistema no nulo de valores de las variables. Análogamente se introducen los conceptos de formas *no positivas* y de formas *definidas negativas*. Las formas no negativas y no positivas se llaman a veces formas de *signo constante*.

Si una forma en n variables $\xi_1, \ldots, \xi_n$ es de forma diagonal

$$\alpha_1\xi_1^2+\ldots+\alpha_s\xi_s^2-\alpha_{s+1}\xi_{s+1}^2-\ldots-\alpha_r\xi_r^2 \quad (\alpha_i>0),$$

es fácil ver que será definida positiva cuando $s=n$, no negativa cuando $s=r\leqslant n$, no positiva cuando $s=0$ y definida negativa cuando $s=0$ y $r=n$. Siendo $0<s<r$ la forma puede tomar valores positivos para unos valores de las variables y valores negativos para otros valores de las variables. Puesto que la constancia de signo de una forma se conserva al realizar una transformación de las variables, se puede decir que son definidas positivas aquellas formas que se reducen a la suma de n cuadrados positivos y que son no negativas aquellas formas que se reducen a la suma de cuadrados positivos solamente, aunque sea en un número menor que el número de las variables; afirmaciones correspondientes tienen lugar para las formas no positivas y definidas negativas. En particular, una forma no negativa es definida positiva sólo cuando es regular.

Si en las formas cuadráticas corrientes se permite que las variables tomen valores tanto reales como complejos, los conceptos introducidos pierden el sentido, ya que en este caso cualquier forma

no nula puede tomar valores no reales. La situación es distinta si se consideran formas hermitianas. La razón de esto estriba en que tanto para los valores reales como complejos de las variables las formas hermitianas toman siempre valores reales. Efectivamente, de las condiciones $\alpha_{ij}=\overline{\alpha}_{ji}$ se deduce que

$$\overline{F}=\sum\overline{\alpha_{ij}\xi_i\overline{\xi}_j}=\sum\overline{\alpha}_{ij}\xi_i\overline{\xi}_j=\sum\alpha_{ji}\xi_j\overline{\xi}_i=F,$$

es decir, se deduce que F es real. Por esto, los conceptos de forma no negativa, definida positiva, etc. se pueden extender directamente al caso de formas cuadráticas hermitianas.

Es fácil establecer si una forma es definida positiva, así como determinar su signatura, reduciéndola a la forma diagonal por el método de Lagrange. Sin embargo, en determinados casos tienen gran interés los criterios directos de definición positiva. Entre estos nos limitaremos a exponer el así llamado criterio de Jacobi.

CRITERIO DE JACOBI *Una forma cuadrática o cuadrática hermitiana en n variables de matriz A es no negativa cuando, y sólo cuando, los coeficientes del polinomio característico de A son de signos alternados. Además, si uno de estos coeficientes se anula, también resultan nulos todos los coeficientes de los términos de grado inferior.*

Efectivamente, por lo visto en el p. 22.1, la forma dada puede ser reducida, mediante una transformación de las variables de matriz unitaria U, a la forma diagonal de matriz diagonal real $A_1=UA\overline{U}'=UAU^{-1}$. Puesto que la matriz A, además de ser congruente, es también semejante de A_1, el polinomio característico de A coincide con el polinomio característico de A_1 y basta demostrar el criterio sólo en el caso de formas diagonales de tipo $\alpha_1\xi_1\overline{\xi}_1+\ldots+\alpha_r\xi_r\overline{\xi}_r$. El polinomio característico de esta última matriz es

$$\varphi(\lambda)=(\lambda-\alpha_1)(\lambda-\alpha_2)\ldots(\lambda-\alpha_r)\lambda^{n-r}=\lambda^n-\sigma_1\lambda^{n-1}+\sigma_2\lambda^{n-2}-\ldots$$

Si todos los α_i son positivos, las fórmulas de Vieta

$$\sigma_i=\sum\alpha_{n_1}\alpha_{n_2}\ldots\alpha_{n_i}\quad(n_1<n_2<\ldots<n_i)$$

muestran que los coeficientes de $\varphi(\lambda)$ poseen la propiedad requerida. Recíprocamente, supongamos que los coeficientes de un polinomio son diferentes de cero y tienen signos alternados y supongamos que las raíces del polinomio son reales. Debemos demostrar que todas sus raíces son positivas. Supongamos, por inducción, que esto ha quedado ya demostrado para todos los polinomios de grado inferior. Entonces $\varphi'(\lambda)$ tendrá $n-1$ raíces positivas (ya que si todas las condiciones indicadas se cumplen para $\varphi(\lambda)$, también se cumplen indudablemente para $\varphi'(\lambda)$). Pero en este caso $\varphi(\lambda$ tiene, según el teorema de Rolle, no menos de $n-1$ raíces positivas y la última raíz n-ésima de $\varphi(\lambda)$ será positiva debido a que el producto de todas las raíces de $\varphi(\lambda)$ es, por hipótesis, positivo

Ejemplos y problemas

1. Hállese la matriz de la transformación lineal de las variables que reduce la forma bilineal

$$\xi_1\eta_2-\xi_2\eta_1+\xi_3\eta_2-\xi_2\eta_3+2\xi_4\eta_1-2\xi_1\eta_4$$

a la forma elemental.

2. Demuéstrese que sobre un espasio unitario real toda forma bilineal antisimétrica real en n variables puede ser reducida, mediante una transformación unitaria real de las variables, a

$$\alpha_1(\xi_1\eta_2-\xi_2\eta_1)+\ldots+\alpha_r(\xi_{2r-1}\eta_{2r}-\xi_{2r}\eta_{2r-1})$$

(compárese con el p. 19.5).

3. Hállense las matrices de las transformaciones ortogonales reales de las variables que reducen a la forma diagonal las formas

a) $\xi_1^2+\xi_1\xi_2+\xi_2^2$;

b) $99\xi_1^2-12\xi_1\xi_2+48\xi_1\xi_3+130\xi_2^2-60\xi_2\xi_3+71\xi_3^2$;

c) $10\xi_1\eta_1+4\xi_1\eta_2+4\xi_2\eta_1+12\xi_1\eta_3+12\xi_3\eta_1-2\xi_2\eta_3-14\xi_3\eta_3-14\xi_3\eta_2+\xi_2\eta_2$

y determínense estas formas elementales.

4. Hállese la matriz de la transformación unitaria que reduce la forma

$$6\xi_1\eta_2-6\xi_2\eta_1+2\xi_1\eta_4-2\xi_4\eta_1+2\xi_2\eta_3-2\xi_3\eta_2-6\xi_3\eta_4+6\xi_4\eta_3$$

a la forma elemental.

5. Hállese la transformación de las variables que reduce la forma

$$\xi_1\xi_n+\xi_2\xi_{n-1}+\ldots+\xi_s\xi_{n-s+1}\qquad\left(s=\left[\frac{n}{2}\right]\right)$$

a la suma de cuadrados.

§ 23. Pares de formas

23.1. Equivalencia de pares de formas. En los puntos anteriores hemos examinado el problema sobre la reducción a la forma elemental de una forma cuadrática. En este parrágrafo será considerado el problema importante acerca de la reducción simultánea de dos formas cuadráticas.

Recordemos que una sucesión de formas bilineales $F_1, \ldots, F_k$ en los mismos sistemas de variables $\xi_1, \ldots, \xi_n$ y $\eta_1, \ldots, \eta_n$ se llama *equivalente* a una sucesión de formas bilineales $G_1, \ldots, G_k$ en las variables $\xi_1', \ldots, \xi_n'$ y $\eta_1', \ldots, \eta_n'$ si mediante unas transformaciones lineales invertibles de las variables

$$[\xi]=[\xi']T,\quad [\eta]=[\eta']S\quad ([\xi]=[\xi_1,\ \ldots,\ \xi_n])$$

las formas de la primera sucesión pueden ser reducidas a las formas correspondientes de la segunda sucesión.

De los resultados del p. 21.2 se deduce que una sucesión de formas bilineales de matrices $A_1, \ldots, A_k$ es equivalente a una sucesión de formas bilineales de matrices $B_1, \ldots, B_k$ cuando,

y sólo cuando, existen unas matrices regulares P y Q con elementos del campo principal tales que

$$PA_jQ = B_j \quad (j = 1, \ldots, k). \tag{1}$$

Dos sucesiones de matrices $A_1, \ldots, A_k$ y $B_1, \ldots, B_k$ se llaman *equivalentes* si están ligadas por las relaciones (1), donde P y Q son unas matrices regulares. Por consiguiente, la equivalencia de unas sucesiones de formas y la equivalencia de las sucesiones de sus matrices son conceptos que tienen el mismo sentido.

Sea λ una variable independiente y sea F y G un par de matrices cuadradas de un mismo orden n. Los factores invariantes de la λ-matriz $\lambda F - G$ (véase el p. 13.2) se llaman *factores invariantes del par* F y G. Si el par F y G es equivalente al par F_1 y G_1, de (1) resulta

$$\lambda F_1 - G_1 = U(\lambda F - G)V'.$$

Por lo visto en el p. 13.3 esto significa que los factores invariantes de las λ-matrices $\lambda F_1 - G_1$ y $\lambda F - G$ coinciden. Es decir, *para la equivalencia de dos pares de matrices es necesario que coincidan los factores invariantes de estos pares.* En el caso general esta condición no es suficiente[1]. Sin embargo, si las primeras matrices de ambos pares son regulares, la coincidencia de los factores invariantes es suficiente para que estos pares sean equivalentes.

TEOREMA 1. *Sean dados dos pares de matrices cuadradas F, G y F_1, G_1 de un mismo orden y sean F y F_1 regulares. Para la equivalencia de los pares F, G y F_1, G_1 es necesario y suficiente que los factores invariantes de la matriz $\lambda F - G$ coincidan con los factores invariantes de la matriz $\lambda F_1 - G_1$.*

La necesidad ha sido ya demostrada y, por ello, estableceremos solamente la suficiencia. Supongamos que los factores invariantes de las matrices $\lambda F - G$ y $\lambda F_1 - G_1$ coinciden. De las relaciones

$$F^{-1}(\lambda F - G) = \lambda E - F^{-1}G,$$
$$F_1^{-1}(\lambda F_1 - G_1) = \lambda E - F_1^{-1}G_1$$

se desprende que también coinciden los factores invariantes de las matrices $\lambda E - F^{-1}G$ y $\lambda E - F_1^{-1}G_1$. Puesto que éstas son las matrices características de $F^{-1}G$ y $F_1^{-1}G_1$, de aquí se deduce (p. 15.3) que $F^{-1}G$ y $F_1^{-1}G_1$ son semejantes, es decir, que existe una matriz regular T que satisface la relación

$$F_1^{-1}G_1 = T^{-1}F^{-1}GT.$$

Tenemos ahora

$$\lambda E - F_1^{-1}G_1 = T^{-1}(\lambda E - F^{-1}G)T = T^{-1}F^{-1}(\lambda F - G)T,$$
$$\lambda F_1 - G_1 = F_1T^{-1}F^{-1}(\lambda F - G)T,$$

[1] En el caso general, además de los factores invariantes, es preciso considerar también los así llamados índices mínimales.

de donde

$$F_1 = UFV' \quad \text{y} \quad G_1 = UGV',$$

donde $U = F_1 T^{-1} F^{-1}$ y $V = T'$. Por consiguiente, los pares F, G y F_1, G_1 son equivalentes.

23.2. Congruencia de pares de formas. Una sucesión de formas bilineales en las variables $\xi_1, \ldots, \xi_n$ y $\eta_1, \ldots, \eta_n$ de matrices $A_1, \ldots, A_k$ se llama *congruente* de una sucesión de formas bilineales de matrices $B_1, \ldots, B_k$ si mediante una misma transformación de ambos sistemas de variables las formas de la primera sucesión pueden ser reducidas a las formas correspondientes de la segunda sucesión.

Indicando por T la matriz de la transformación de las variables, tendremos

$$B_j = TA_jT' \quad (j = 1, 2, \ldots, k). \tag{2}$$

Dos sucesiones $A_1, \ldots, A_k$ y $B_1, \ldots, B_k$ de matrices arbitrarias se llaman *congruentes* si existe una matriz regular T que cumple las condiciones (2).

Análogamente se dice que una sucesión de formas cuadráticas o de formas cuadráticas hermitianas es congruente de otra sucesión de formas, si mediante una transformación adecuada invertible de las variables las formas de la primera sucesión pueden ser reducidas simultáneamente a las formas correspondientes de la segunda sucesión.

Es obvio que la condición (2) es necesaria y suficiente para la congruencia de las formas cuadráticas de matrices $A_1, \ldots, A_k$ y $B_1, \ldots, B_k$. Para las formas hermitianas esta condición debe ser sustituida por la siguiente

$$B_j = TA_j\bar{T}' \quad (j = 1, 2, \ldots, k).$$

El problema general que hemos planteado acerca de la congruencia de sucesiones de formas bilineales es muy complejo ya para un par de formas. En la segunda mitad del siglo pasado Weierstrass obtuvo las condiciones necesarias y suficientes de congruencia en el caso en que ambas formas del par son simétricas y, en particular, si las formas son cuadráticas y una forma del par es regular. Estas condiciones serán expuestas al concluir este punto. El caso general de congruencia de pares de formas cuadráticas fue examinado por Kronecker. Debido a que las condiciones de Kronecker son un tanto voluminosas, suelen exponerse en tratados más especiales. Tanto las primeras condiciones como las segundas se refieren a formas sobre el cuerpo de los números complejos. El caso de cuerpos conmutativos de otros tipos fue analizado por Dixon y otros autores.

Consideraremos primero el caso más importante de un par de formas cuadráticas reales cuando una de ellas es definida positiva.

TEOREMA 2. *Todo par de formas cuadráticas reales en n variables tal que una de ellas es definida positiva puede ser reducido, mediante una adecuada transformación real de las variables, al par de formas de tipo*

$$\xi_1^2+\xi_2^2+\ldots+\xi_n^2 \quad \text{y} \quad \alpha_1\xi_1^2+\alpha_2\xi_2^2+\ldots+\alpha_n\xi_n^2. \tag{3}$$

Los números $\alpha_1, \ldots, \alpha_n$ se determinan unívocamente, salvo el orden en que siguen, por las formas iniciales y no dependen del método de reducción.

En efecto, sean $F(\xi)$ y $G(\xi)$ las formas cuadráticas dadas. Puesto que la primera es definida positiva, es posible reducirla, mediante una adecuada transformación lineal de las variables, a la suma de los cuadrados de las variables. Así obtenemos el par de formas de tipo

$$\xi_1^2+\xi_2^2+\ldots+\xi_n^2 \quad \text{y} \quad \sum a_{ij}\xi_i\xi_j.$$

Ahora buscamos una transformación de las variables de matriz *ortogonal* real U que reduzca a la forma diagonal la segunda forma. Como en este caso la matriz de la primera forma en las variables nuevas será igual a

$$UEU'=UU'=E,$$

es decir, la primera forma conserva su forma unitaria, resulta que después de la transformación señalada las formas dadas serán del tipo requerido (3).

Supongamos, finalmente, que mediante una transformación de las variables las formas F y G se reducen al tipo (3) y mediante otra transformación se reducen a

$$\xi_1^2+\xi_2^2+\ldots+\xi_n^2 \text{ y } \beta_1\xi_1^2+\beta_2\xi_2^2+\ldots+\beta_n\xi_n^2. \tag{4}$$

Existe entonces una transformación de las variables de matriz T que reduce el par (3) en el par de formas (4) y, por consiguiente, las matrices de estas formas estarán ligadas por las relaciones

$$E=TET' \quad \text{y} \quad B=TAT', \tag{5}$$

donde A es la matriz diagonal de elementos $\alpha_1, \ldots, \alpha_n$ y B es la matriz diagonal de elementos $\beta_1, \ldots, \beta_n$ a lo largo de la diagonal principal. De la primera de las relaciones (5) resulta $TT'=E$, de donde $B=TAT^{-1}$, es decir, las matrices A y B son semejantes y los números característicos $\alpha_1, \ldots, \alpha_n$ de la primera deben coincidir con los números característicos $\beta_1, \ldots, \beta_n$ de la segunda.

Para las formas cuadráticas hermitianas tiene lugar un teorema totalmente análogo.

TEOREMA 2a. *Todo par de formas cuadráticas hermitianas tal que la primera es definida positiva se puede reducir, mediante una adecuada transformación lineal compleja de las variables, en el par de*

formas de tipo

$$\xi_1\overline{\xi}_1+\xi_2\overline{\xi}_2+\ldots+\xi_n\overline{\xi}_n \quad \text{y} \quad \alpha_1\xi_1\overline{\xi}_1+\alpha_2\xi_2\overline{\xi}_2+\ldots+\alpha_n\xi_n\overline{\xi}_n,$$

donde los números $\alpha_1, \ldots, \alpha_n$ *se determinan unívocamente por las formas iniciales y no dependen del método de reducción.*

La demostración es análoga a la anterior.

Pasando a considerar el caso general observemos, ante todo, que la congruencia de un par de matrices implica indudablemente la equivalencia de las mismas. Inesperadamente resulta que bajo ciertas condiciones es válida también la afirmación recíproca.

TEOREMA 3. *Supongamos que en dos pares de matrices cuadradas* F, G y F_1, G_1 *las primeras matrices* F *y* F_1 *son ambas o bien simétricas o bien antisimétricas y que las segundas matrices* G *y* G_1 *son ambas también o bien simétricas o bien antisimétricas. Entonces, sobre el cuerpo de los números complejos, de la equivalencia de los pares señalados se deduce la congruencia de los mismos.*

Existen, por hipótesis, unas matrices regulares U y V tales que

$$F_1=UFV' \quad \text{y} \quad G_1=UGV'. \tag{6}$$

Pasando a las matrices transpuestas, obtenemos de aquí que $F_1'=VF'U'$ y $G_1'=VG'U'$. Puesto que las matrices F y F_1 son simétricas o antisimétricas de aquí resulta

$$F_1=VFU'; \tag{7}$$

una igualdad análoga tenemos para G_1. Comparando (7) y (6) encontramos

$$UFV'=VFU' \text{ y } V^{-1}U\cdot F=F\cdot(V^{-1}U)'. \tag{8}$$

Tomemos $V^{-1}U=T$. La segunda de las igualdades (8) da

$$\begin{aligned} TF&=FT',\\ T^2F&=FT'^2,\\ &\ldots\ldots\\ T^kF&=FT'^k, \end{aligned}$$

de donde

$$(\alpha_0E+\alpha_1T+\ldots+\alpha_kT^k)F=F(\alpha_0E+\alpha_1T'+\ldots+\alpha_kT'^k),$$

donde $\alpha_0, \alpha_1, \ldots, \alpha_k$ son unos números arbitrarios. En el capítulo IV (p. 16.3) hemos demostrado que los números $\alpha_0, \ldots, \alpha_k$ se pueden escoger de manera que el polinomio

$$\varphi(T)=\alpha_0E+\alpha_1T+\ldots+\alpha_kT^k$$

sea raíz cuadrada de T, es decir, que $\varphi(T)\varphi(T)=T$. Tomemos $P=V\varphi(T)$; entonces

$$PFP'=V\varphi(T)F\varphi(T')V'=V\varphi(T)\varphi(T)FV'=VTFV'=UFV',$$

es decir,

$$PFP' = F_1.$$

Repitiendo estos mismos razonamientos para la matriz G, encontramos

$$PGP' = G_1.$$

Por consiguiente, los pares F, G y F_1, G_1 son congruentes que es lo que se quería demostrar.

Los teoremas 1 y 3 permiten enunciar la siguiente condición de congruencia de un par de matrices:

Sean dados dos pares de matrices F, G y F_1, G_1. Si F y F_1 son regulares y son ambas o bien simétricas o bien antisimétricas y si G y G_1 también son ambas o bien simétricas o bien antisimétricas, la condición necesaria y suficiente de congruencia de los pares F, G y F_1, G_1 sobre el cuerpo de los números complejos es la coincidencia de los factores invariantes de las matrices $\lambda F - G$ y $\lambda F_1 - G_1$.

Efectivamente, si los factores invariantes de las matrices $\lambda F - G$ y $\lambda F_1 - G_1$ coinciden, los pares F, G y F_1, G_1 son equivalentes en virtud del teorema 1. Pero estos pares serán entonces, debido al teorema 3, también congruentes. Recíprocamente, si F, G y F_1, G_1 son congruentes, son desde luego equivalentes y, por consiguiente, los factores invariantes de la matriz $\lambda F - G$ coinciden con los factores invariantes de la matriz $\lambda F_1 - G_1$.

Las aplicaciones del teorema 3 a la determinación directa de las formas elementales de pares de formas complejas serán consideradas más tarde, en el cap. VII. Aplicando este teorema obtendremos en el punto siguiente la solución del problema de congruencia de las formas bilineales no simétricas.

Observemos también que si hasta el teorema 3 las formas cuadráticas reales y las formas hermitianas se comportaban igualmente, el teorema 3 deja ya de ser válido para las formas hermitianas. El ejemplo correspondiente será expuesto al principio del § 28.

23.3. Congruencia de formas bilineales no simétricas. Como ya hemos señalado anteriormente, los resultados del punto precedente permiten enunciar las condiciones necesarias y suficientes para la congruencia de cualesquiera formas bilineales complejas regulares complementando con ello esencialmente los resultados del p. 21.3.

TEOREMA 4. *Para que unas formas bilineales complejas regulares de matrices G y G_1 sean congruentes es necesario y suficiente que coincidan los divisores elementales de las λ-matrices $\lambda G - G'$ y $\lambda G_1 - G_1'$.*

DEMOSTRACIÓN. Supongamos que las formas son congruentes. Las matrices G y G_1 están ligadas entonces por las relaciones

$$G_1 = UGU'.$$

De aquí

$$G_1' = UG'U'.$$

Por consiguiente, el par de matrices G, G' es congruente del par de matrices G_1, G_1' y los divisores elementales de la matriz $\lambda G - G'$ coinciden con los divisores elementales de la matriz $\lambda G_1 - G_1'$.

Supongamos, recíprocamente, que los divisores elementales de las matrices $\lambda G - G'$ y $\lambda G_1 - G_1'$ coinciden. Entonces el par G, G' es equivalente al par G_1, G_1', en virtud del teorema 1, por lo tanto

$$G_1 = UGV' \quad \text{y} \quad G_1' = UG'V'. \tag{9}$$

Tomemos

$$G + G' = S, \quad G - G' = T, \quad G_1 + G_1' = S_1 \quad \text{y} \quad G_1 - G_1' = T_1.$$

De (9) resulta que

$$S_1 = USV' \quad \text{y} \quad T_1 = UTV'$$

es decir, el par S, T es equivalente al par S_1, T_1. Puesto que las matrices S y S_1 son simétricas y las matrices T y T_1 son antisimétricas, los pares S, T y S_1, T_1 son, en virtud del teorema 3, congruentes, es decir,

$$S_1 = PSP' \quad \text{y} \quad T_1 = PTP'. \tag{10}$$

Pero $G = \frac{S+T}{2}$ y $G_1 = \frac{S_1+T_1}{2}$ y, por esto, de (10) se desprende que

$$G_1 = PGP'.$$

Por consiguiente, las matrices G y G_1 son congruentes.

El teorema 4 muestra que las formas bilineales regulares sobre el cuerpo de los números complejos se determinan, salvo un isomorfismo, por los divisores elementales de la matriz $\lambda G - G'$. Luego, para resolver el problema de la clasificación de estas formas es suficiente indicar los sistemas de expresiones de tipo $(\lambda - \alpha)^m$ que pueden servir como divisores elementales de las matrices de tipo $\lambda G - G'$. Es fácil ver que estos sistemas no pueden ser arbitrarios. En efecto, supongamos que $(\lambda - \alpha)^m$ figura k veces en el sistema de divisores elementales de una matriz $\lambda G - G'$. Entonces $(\lambda - \alpha)^m$ figurará también k veces en el sistema de divisores elementales de la matriz transpuesta $\lambda G' - G$. Pero la matriz $\lambda G' - G$ es equivalente a la matriz $G'^{-1}(\lambda G' - G) = \lambda E - G'^{-1}G$; por consiguiente, $(\lambda - \alpha)^m$ figura k veces en el sistema de divisores elementales de la matriz $\lambda E - G'^{-1}G$. El teorema sobre los divisores elementales de una función (p. 16.4) afirma que todo divisor elemental $(\lambda - \alpha)^m$ de la matriz $\lambda E - G'^{-1}G$ se transforma en el divisor elemental $(\lambda - \alpha^{-1})^m$ de la matriz $\lambda E - (G'^{-1}G)^{-1}$. Puesto que la última matriz es equivalente a $\lambda G - G'$, la expresión $(\lambda - \alpha^{-1})^m$ figura k veces en el sistema de divisores elementales de la matriz $\lambda G - G'$. Por

consiguiente, *si* $(\lambda-\alpha)^m$ *es un divisor elemental de mutiplicidad k de la matriz* $\lambda G-G'$, *resulta que* $(\lambda-\alpha^{-1})^m$ *también será un divisor elemental de multiplicidad k de la misma.* Siendo $\alpha=\pm 1$, esta condición es trivial. Con razonamientos adicionales se puede demostrar que las matrices $\lambda G-G'$ contienen todo divisor de tipo $(\lambda+1)^{2m+1}$ necesariamente un número par de veces, mientras que los divisores elementales de tipo $(\lambda-1)^{2m+1}$ y $(\lambda\pm 1)^{2m}$ pueden figurar en estas matrices en combinaciones arbitrarias. Las condiciones expuestas, para α, tanto iguales como diferentes de ± 1, además de ser necesarias, son también suficientes para que un sistema de expresiones de tipo $(\lambda-\alpha)^m$ sea el sistema de divisores elementales de una matriz $\lambda G-G'$.

Ejemplos y problemas

1. Redúzcanse a la forma elemental, mediante una transformación real de las variables, los pares de formas

a) $x^2+2xy+2y^2$ y $2x^2-xy$;

b) $2x^2+2xy+2xz+2y^2+2yz+2z^2$ y $-9x^2+36xy+18xz+15y^2+18yz+18z^2$.

2. Todo par compuesto por formas bilineales reales, una de las cuales es simétrica definida positiva y la otra es antisimétrica, puede ser reducido, mediante una transformación real de las variables, a la forma

$$\xi_1\eta_1+\xi_2\eta_2+\ldots+\xi_n\eta_n \text{ y } \alpha_1(\xi_1\eta_2-\xi_2\eta_1)+\ldots+\alpha_r(\xi_{2r-1}\eta_{2r}-\xi_{2r}\eta_{2r-1}).$$

3. Demuéstrese que el par de formas de matrices

$$\begin{bmatrix}-2 & -6\\ 1 & 3\end{bmatrix} \text{ y } \begin{bmatrix}-6 & 6\\ 1 & -1\end{bmatrix}$$

es equivalente al par de formas de matrices

$$\begin{bmatrix}2 & -3\\ -2 & 3\end{bmatrix} \text{ y } \begin{bmatrix}-10 & 5\\ 2 & -1\end{bmatrix}.$$

Para la demostración generalícese el teorema 1 del p. 23.1 al caso de pares A, B tales que $|\lambda A+\mu B|\not\equiv 0$, considerando para ello los factores invariantes que son polinomios homogéneos en λ y μ.

4. Demuéstrense las últimas afirmaciones del p. 23.3.

§ 24. Funciones bilineales

La teoría de formas bilineales puede ser interpretada geométricamente como la teoría de funciones bilineales en espacios lineales. Esta interpretación permite al mismo tiempo comprender con más profundidad los resultados principales de la teoría de formas; el parágrafo presente está dedicado a la exposición de la misma.

24.1. Definiciones principales. Se dice que sobre un espacio lineal $\mathfrak{L}$ se ha definido una función $\varphi(x, y)$ de dos vectores variables x e y, si a todo par de vectores del espacio $\mathfrak{L}$ se pone en

correspondencia un determinado elemento $\varphi(x, y)$ del campo principal K del espacio considerado. La función $\varphi(x, y)$ se llama *bilineal* si es lineal respecto a cada una de las variables por separado, es decir, si satisface las identidades

$$\varphi(\alpha_1 x_1 + \alpha_2 x_2, y) = \alpha_1 \varphi(x_1, y) + \alpha_2 \varphi(x_2, y), \quad (1)$$

$$\varphi(x, \beta_1 y_1 + \beta_2 y_2) = \beta_1 \varphi(x, y_1) + \beta_2 \varphi(x, y_2). \quad (2)$$

Una función bilineal sobre un espacio complejo lineal se llama *bilineal hermitiana*, si es lineal respecto a la primera variable y antilineal respecto a la segunda, es decir, si satisface la relación (1) y la relación

$$\varphi(x, \beta_1 y_1 + \beta_2 y_2) = \overline{\beta}_1 \varphi(x, y_1) + \overline{\beta}_2 \varphi(x, y_2). \quad (3)$$

De (1), (2) y (3) se deduce directamente la regla general distributiva

$$\varphi(x_1 + x_2 + \ldots + x_m, y_1 + y_2 + \ldots + y_s) = \sum \varphi(x_i, y_j) \quad (4)$$

que es válida tanto para las formas bilineales como para las bilineales hermitianas.

Tomemos en $\mathfrak{L}$ una base arbitraria $a_1, \ldots, a_n$ y sea

$$x = \xi_1 a_1 + \xi_2 a_2 + \ldots + \xi_n a_n,$$
$$y = \eta_1 a_1 + \eta_2 a_2 + \ldots + \eta_n a_n.$$

De las fórmulas (2) y (4) resulta

$$\varphi(x, y) = \sum \varphi(a_i, a_j) \xi_i \eta_j = \sum \alpha_{ij} \xi_i \xi_j \quad (\alpha_{ij} = \varphi(a_i, a_j)). \quad (5)$$

La matriz $A = \|\alpha_{ij}\|$ se llama *matriz de la función* $\varphi(x, y)$ en la base señalada. Es obvio que conociendo la matriz A conocemos también la función $\varphi(x, y)$, ya que la fórmula (5) permite calcular los valores de $\varphi(x, y)$ cualquiera que sea el par de vectores x, y.

La correspondencia entre las matrices y las formas bilineales es biyectiva, ya que cualquiera que sea la matriz A la función $\varphi(x, y)$ calculada por la fórmula (5) será bilineal de matriz A.

Indicando por $[x]$ e $[y]$ las filas coordenadas de los vectores x e y, podemos representar la fórmula (5) en la forma matricial

$$\varphi(x, y) = [x] \; A \; [y]', \quad (6)$$

que permite deducir directamente la siguiente *regla de transformación* de las matrices de funciones bilineales: *si en una base la matriz de una función bilineal es A, la matriz de la función en una base nueva será*

$$A_1 = TAT', \quad (7)$$

donde T es la matriz del cambio de la base antigua por la nueva.

Efectivamente, aplicando la fórmula (6) en las bases antigua y nueva, tenemos

$$\varphi(x, y) = [x]\, A\, [y]' = [x]_1\, TAT'\, [y]'_1 = [x]_1\, A_1\, [y]'_1,$$

donde A_1, $[x]_1$ e $[y]_1$ son, respectivamente, la matriz de la función $\varphi(x, y)$ y las filas coordenadas de los vectores x e y en la base nueva. Por esto, $A_1 = TAT'$.

Considerando en la fórmula (5) las coordenadas de los vectores x e y como variables independientes, vemos que *el valor de una función bilineal en toda base se expresa mediante una forma bilineal cuya matriz coincide con la matriz de la función en esta base.*

Según (7), el paso a una base nueva implica la sustitución de la forma bilineal por la correspondiente forma congruente, debido a lo cual *las formas bilineales congruentes pueden ser consideradas como formas bilineales de una misma función bilineal, pero calculadas en diferentes bases.*

Una función bilineal $\varphi(x, y)$ se llama *simétrica* si

$$\varphi(x, y) = \varphi(y, x)$$

y *antisimétrica* si

$$\varphi(x, y) = -\varphi(x, y).$$

Es evidente que *las funciones bilineales simétricas y antisimétricas son aquellas funciones cuyas formas bilineales son, respectivamente, simétricas y antisimétricas.*

Si $\varphi(x, y)$ es una función bilineal arbitraria, las funciones

$$\varphi_1(x, y) = \frac{1}{2}[\varphi(x, y) + \varphi(y, x)] \text{ y}$$

$$\varphi_2(x, y) = \frac{1}{2}[\varphi(x, y) - \varphi(y, x)] \qquad (8)$$

serán bilineales y, respectivamente, simétrica y antisimétrica. Como de (8) se deduce que

$$\varphi(x, y) = \varphi_1(x, y) + \varphi_2(x, y)$$

resulta que toda función bilineal puede ser representada como la suma de funciones simétrica y antisimétrica, con la particularidad de que, como es fácil de ver, está representación es unívoca.

Las funciones de tipo $\psi(x) = \varphi(x, x)$, donde $\varphi(x, y)$ es una función bilineal, se llaman *cuadráticas*. Por consiguiente, las funciones cuadráticas son aquellas funciones de un vector variable que aparecen al identificar los vectores variables en las funciones bilineales. Pero si una función cuadrática $\psi(x)$ aparece de la forma explicada de una función bilineal $\varphi(x, y)$, la función $\psi(x)$ aparece también al identificar las variables en la función bilineal simétrica $\varphi_1(x, y)$:

$$\varphi_1(x, x) = \frac{1}{2}[\varphi(x, x) + \varphi(x, x)] = \varphi(x, x).$$

Por esto al considerar las funciones cuadráticas se puede aceptar que son generadas por funciones simétricas solamente.

Por otra parte, si $\varphi(x, y)$ es simétrica y $\psi(x) = \varphi(x, x)$, se tiene

$$\psi(x+y) = \varphi(x+y, x+y) = \varphi(x, x) + 2\varphi(x, y) + \varphi(y, y)$$

o

$$\varphi(x, y) = \frac{1}{2}[\psi(x+y) - \psi(x) - \psi(y)],$$

es decir, *toda función cuadrática aparece de una función simétrico bilineal y sólo una, que se llama polar de la función cuadrática correspondiente.*

Sea $a_1, \ldots, a_n$ una base del espacio y sea $\psi(x) = \varphi(x, x)$, donde $\varphi(x, y)$ es una función bilineal simétrica. Entonces para $x = \xi_1 a_1 + \ldots + \xi_n a_n$ tenemos

$$\psi(x) = \varphi(x, x) = \sum \varphi(a_i, a_j)\, \xi_i \xi_j,$$

es decir, *el valor de una función cuadrática se expresa mediante una forma cuadrática en las coordenadas de un vector variable, cuya matriz coincide con la matriz de la correspondiente forma bilineal polar.*

Razonando análogamente, obtenemos para los valores de una función bilineal *hermitiana* $\varphi(x, y)$ en lugar de la fórmula (5) la fórmula

$$\varphi(x, y) = \sum \varphi(a_i, a_j)\, \xi_i \overline{\eta}_j,$$

es decir, una forma bilineal hermitiana. Al pasar a una base nueva la matriz de una función hermitiana varía según la ley

$$A_1 = T A \overline{T}'.$$

Una función hermitiana $\varphi(x, y)$ se llama simétrica si

$$\varphi(x, y) = \overline{\varphi(y, x)}.$$

Los valores de las funciones hermitianas simétricas se expresan mediante formas bilineales hermitianas simétricas.

Las funciones de un vector variable que surgen al identificar los vectores variables en las funciones bilineales hermitianas *simétricas* se llaman *funciones cuadráticas hermitianas*. La relación entre las funciones cuadráticas hermitianas y las coorrespondientes funciones bilineales simétricas hermitianas viene dada por las fórmulas

$$\left.\begin{aligned} &\psi(x+iy) = \varphi(x+iy, x+iy) = \\ &= \varphi(x, x) + i[\overline{\varphi(x, y)} - \varphi(x, y)] + \varphi(y, y), \\ &\psi(x+y) = \varphi(x, x) + \overline{\varphi(x, y)} + \varphi(x, y) + \varphi(y, y), \\ &2\varphi(x, y) = \psi(x+y) + i\psi(x+iy) - \\ &\qquad - (1+i)[\psi(x) + \psi(y)]. \end{aligned}\right\} \qquad (9)$$

Calculando el valor de una función $\psi(x)$ cuadrática hermitiana para el vector $x = \xi_1 a_1 + \ldots + \xi_n a_n$, donde $a_1, \ldots, a_n$ es una base del espacio, obtenemos

$$\psi(x) = \sum \varphi(a_i, a_j)\, \xi_i \overline{\xi}_j,$$

es decir, *el valor de una función cuadrática hermitiana se expresa mediante una forma cuadrática hermitiana en las coordenadas del vector variable, cuya matriz coincide con la matriz de la correspondiente función bilineal simétrica hermitiana calculada mediante la fórmula* (9).

Indiquemos también que las formas cuadráticas de signo constante, consideradas en el p. 22.4, corresponden a aquellas funciones cuadráticas en el caso de un espacio real y a aquellas funciones cuadráticas hermitianas en el caso de un espacio complejo cuyos valores no cambian de signo.

24.2. Espacios de métrica bilineal. Como hemos señalado en el p. 17.1, un espacio lineal complejo o real se llama unitario si en él está definida una función de dos vectores variables, llamándose producto escalar de estos vectores los valores de la misma. Los axiomas que hemos indicado en esa ocasión significan simplemente que el producto escalar es una función bilineal hermitiana definida positiva. Por esto, el estudio de las propiedades de los espacios unitarios no es otra cosa que el estudio, desde un punto de vista especial, de las propiedades de las funciones bilineales definidas positivas.

Por analogía con esto diremos que un espacio lineal $\mathfrak{L}$ es *bilineal métrico*, si en él está definida una función bilineal, cuyos valores se llamarán productos escalares de los vectores y se indicarán por (a, b). En el caso en el que el producto escalar sea una función bilineal hermitiana simétrica diremos que el espacio está provisto de una métrica bilineal hermitiana.

Si la matriz del producto escalar es regular, el espacio se llama *no degenerado*. En el caso contrario se dice que el espacio es *degenerado*.

La matriz

$$G = \begin{bmatrix} (a_1, a_1) & \ldots & (a_1, a_m) \\ (a_2, a_1) & \ldots & (a_2, a_m) \\ \ldots & \ldots & \ldots \\ (a_m, a_1) & \ldots & (a_m, a_m) \end{bmatrix},$$

formada por los productos escalares de los vectores $a_1, \ldots, a_m$ de un espacio bilineal métrico $\mathfrak{L}$, se llama *matriz de Gram* del sistema $a_1, \ldots, a_m$. La matriz de Gram correspondiente a unos vectores $a_1, \ldots, a_n$ que constituyen una *base* del espacio $\mathfrak{L}$, se llama simplemente matriz de la función bilineal principal $\varphi(x, y) = (x, y)$

calculada en esta base. De aquí se deduce que las matrices de Gram correspondientes a diferentes bases del espacio son congruentes y, por consiguiente, son del mismo rango.

En la teoría de los espacios unitarios desempeña un papel importante el concepto de ortogonalidad de vectores. Este concepto se extiende directamente al caso de espacios bilineales métricos.

Sea $\mathfrak{L}$ un espacio bilineal métrico corriente o hermitiano. Se dice que un vector a de $\mathfrak{L}$ es ortogonal a un vector b, si $(a, b)=0$. Si a es ortogonal a b, ello no significa todavía que b sea ortogonal a a, ya que en el caso general $(a, b) \neq (b, a)$. Por esto a veces se dice, para una mayor claridad que *a es ortogonal a b a la izquierda y que b es ortogonal a la derecha a a.* De las leyes distributivas (1) y (2) se deduce que siendo a ortogonal a varios vectores $a_1, \ldots, a_m$, el vector a será ortogonal a cualquier combinación lineal de los mismos. De aquí resulta, a su vez, que *el conjunto de los vectores de un espacio $\mathfrak{L}$ ortogonales todos ellos a la derecha a los vectores de un sistema $\mathfrak{M}$ constituye un subespacio lineal del espacio $\mathfrak{L}$.* Indicaremos este subespacio por $\mathfrak{M}^{\perp}$. El conjunto de los vectores de $\mathfrak{L}$ ortogonales a la izquierda a $\mathfrak{M}$ también es un subespacio lineal que indicaremos ${}^{\perp}\mathfrak{M}$.

Diremos que un vector x es *isótropo a la izquierda* en el espacio $\mathfrak{L}$, si es ortogonal a la izquierda a todos los vectores de $\mathfrak{L}$. El subespacio ${}^{\perp}\mathfrak{L}$, formado por todos los vectores isótropos a la izquierda de $\mathfrak{L}$, se llama *subespacio isótropo a la izquierda* de $\mathfrak{L}$. Análogamente se definen los *vectores isótropos a la derecha y el subespacio isótropo a la derecha* $\mathfrak{L}^{\perp}$.

TEOREMA 1. *Las dimensiones de los subespacios isótropos a la izquierda y a la derecha son iguales y coinciden con el defecto de la matriz de la forma métrica (x, y) calculada en una base cualquiera. Por consiguiente, la diferencia entre la dimensión del espacio y la dimensión de los subespacios isótropos el igual al rango de la forma bilineal métrica; un espacio bilineal métrico es no degenerado cuando, y sólo cuando, no contiene vectores isótropos no nulos.*

Tomemos para la demostración una base de $\mathfrak{L}$. Entonces los vectores x isótropos a la izquierda deben verificar para cualquier y de $\mathfrak{L}$ la relación

$$(x, y) = [x]\, A\, [y]' = 0, \tag{10}$$

donde A es la matriz de Gram de la base tomada. Pero de (10) se deduce que

$$[x]\, A = O$$

(aquí O es la fila nula), es decir los vectores isótropos a la izquierda forman el núcleo de la aplicación lineal de matriz A (p. 10.1) y la dimensión del núcleo de una aplicación lineal es igual al defecto de la matriz de la aplicación. Análogamente comprobamos

que la dimensión del subespacio isótropo a la derecha es igual a la dimensión del núcleo de la aplicación lineal de matriz transpuesta A' y, por consiguiente, coincide con el defecto de la matriz A que es lo que se quería demostrar.

Todo subespacio lineal $\mathfrak{A}$ de $\mathfrak{L}$ puede ser considerado por sí sólo como un espacio bilineal métrico hermitiano o corriente, respectivamente, respecto al mismo producto escalar que opera en $\mathfrak{L}$. En el caso general, si $\mathfrak{L}$ es no degenerado, ello no implica todavía que sus subespacios sean no degenerados y recíprocamente, si un subespacio $\mathfrak{A}$ es no degenerado, esto no implica que todo el espacio $\mathfrak{L}$ sea no degenerado.

TEOREMA 2. *Si $\mathfrak{A}$ es un subespacio no degenerado del espacio $\mathfrak{L}$, para $\mathfrak{L}$ tienen lugar las descomposiciones directas*

$$\mathfrak{L} = \mathfrak{A} \dotplus \mathfrak{A}^{\perp} = {}^{\perp}\mathfrak{A} \dotplus \mathfrak{A}. \tag{11}$$

Recíprocamente, si al menos una de las descomposiciones (11) *es válida, el subespacio $\mathfrak{A}$ es no degenerado.*

Efectivamente, la intersección $\mathfrak{A} \cap \mathfrak{A}^{\perp}$ es el subespacio isótropo a la derecha de $\mathfrak{A}$. Como $\mathfrak{A}$ es no degenerado, tenemos $\mathfrak{A} \cap \mathfrak{A}^{\perp} = o$ y, por consiguiente, la suma $\mathfrak{A} + \mathfrak{A}^{\perp}$ es directa.

Sea ahora c un vector cualquiera de $\mathfrak{L}$. Tomemos

$$(a_j, c) = \gamma_j \quad (j = 1, \ldots, m),$$

donde $a_1, \ldots, a_m$ es una base de $\mathfrak{A}$. El sistema auxiliar de ecuaciones

$$(a_j, a_1)\xi_1 + (a_j, a_2)\xi_2 + \ldots + (a_j, a_m)\xi_m = \gamma_j \quad (j = 1, \ldots, m)$$

puede ser resuelto respecto a $\xi_1, \ldots, \xi_m$, ya que su determinante es el determinante de la matriz de Gram del sistema $a_1, \ldots, a_m$ y es diferente de cero debido a que $\mathfrak{A}$ es no degenerado. El vector $a = \xi_1 a_1 + \ldots + \xi_m a_m$ pertenece a $\mathfrak{A}$ y, además,

$$(a_j, c - a) = 0 \quad (j = 1, \ldots, m),$$

es decir, $c - a \in \mathfrak{A}^{\perp}$. Puesto que para todo c es válida la descomposición

$$c = a + (c - a) \quad (a \in \mathfrak{A} \text{ y } c - a \in \mathfrak{A}^{\perp}),$$

resulta que $\mathfrak{L}$ es la suma de $\mathfrak{A}$ y $\mathfrak{A}^{\perp}$ que es lo que se quería demostrar.

Unos espacios bilineales métricos sobre un mismo cuerpo conmutativo de coeficientes se llaman *isomorfos*, cuando entre sus elementos se puede establecer una correspondencia biyectiva que transforma una suma de vectores en la suma correspondiente, el producto de un número por un vector en el producto del mismo número por el vector correspondiente y el producto escalar de un par de vectores en el producto escalar del par de vectores correspondientes.

De la última condición se ve, en particular, que en una correspondencia isomorfa las matrices de Gram de los sistemas correspondientes de vectores coinciden. Viceversa, si en dos espacios bilineales métricos sobre un mismo cuerpo conmutativo existen bases con las mismas matrices, los espacios son isomorfos.

Para los espacios lineales corrientes y unitarios hemos demostrado antes que existe un espacio único, salvo un isomorfismo, de dimensión n. Para los espacios bilineales métricos la situación es más compleja.

Demostremos, ante todo, la siguiente afirmación casi evidente:

TEOREMA 3. *Unos espacios bilineales métricos sobre un mismo cuerpo conmutativo son isomorfos cuando, y sólo cuando, las matrices de Gram unas bases arbitrarias, escogidas en estos espacios, son congruentes.*

La necesidad es clara, ya que las matrices de Gram de todas las bases de un espacio dado son congruentes y las matrices de Gram de las bases correspondientes de espacios isomorfos coinciden. Por otro lado, si A y B son las matrices de Gram de unas bases de unos espacios $\mathfrak{L}$ y $\mathfrak{L}_1$ y A y B son congruentes, en $\mathfrak{L}$ existe también una base de matriz B, de acuerdo con los resultados del punto anterior.

El teorema 3 significa que el problema de la clasificación de espacios bilineales métricos no isomorfos es idéntico al problema de la clasificación, salvo una congruencia, de las formas bilineales; este problema ha sido ya examinado en el p. 23.3. Nos limitaremos aquí a enunciar, en términos de la teoría de espacios bilineales métricos, algunos corolarios de los resultados del punto mencionado.

Los espacios bilineales métricos reales no degenerados de métrica simétrica $(x, y) = (y, x)$ se llaman *seudoeuclídeos*.

La matriz de Gram A de una base arbitraria de un espacio seudoeuclídeo es simétrica real. Según el teorema 1 del p. 21.3, la matriz A es congruente a una matriz diagonal con los números $+1$ o -1 en la diagonal principal. En otras palabras, en todo espacio seudoeuclídeo de dimensión n existe una base en la que el producto escalar de los vectores de coordenadas $\xi_1, \ldots, \xi_n$ y $\eta_1, \ldots, \eta_n$ se expresa mediante la forma

$$(x, y) = \xi_1\eta_1 + \ldots + \xi_s\eta_s - \xi_{s+1}\eta_{s+1} - \ldots - \xi_n\eta_n.$$

El número $\sigma = s - (n - s)$ se llama *signatura* del espacio.

Por consiguiente, *los espacios seudoeuclídeos se determinan, salvo un isomorfismo, por su dimensión y su signatura.* Además, para cualquier $n > 0$ y cualquier s $(0 \leqslant s \leqslant n)$ existe un espacio seudoeuclídeo de dimensión n y de signatura $s - (n - s)$.

Los espacios bilineales métricos no degenerados de métrica antisimétrica $(x, y) = -(y, x)$ se llaman *simplíciales*.

La matriz de Gram de una base de un espacio simplicial es antisimétrica y por esto, debido al p. 15.5, es congruente a una

matriz celular diagonal con células de tipo $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ a lo largo de la diagonal principal. Por consiguiente, *la dimensión de un espacio simplicial es siempre un número par y para todo n existe, salvo un isomorfismo, un espacio simplicial único de dimensión 2n. En todo espacio simplicial de dimensión 2n existe una base en la que el producto escalar es de la forma*

$$(x, y) = \xi_1\eta_2 - \xi_2\eta_1 + \ldots + \xi_{2n-1}\eta_{2n} - \xi_{2n}\eta_{2n-1},$$

donde $\xi_1, \ldots, \xi_{2n}$ *y* $\eta_1, \ldots, \eta_{2n}$ *son las coordenadas de los vectores* x *e* y.

Un espacio bilineal métrico complejo no degenerado de métrica simétrica se llama *euclídeo complejo*. Puesto que toda forma bilineal simétrica regular se puede reducir sobre el cuerpo de los números complejos a la forma de matriz unidad, para toda dimensión n existe, salvo un isomorfismo, un espacio euclídeo complejo único. *En todo espacio euclídeo de dimensión n existe una base en la que el producto escalar es de la forma*

$$(x, y) = \xi_1\eta_1 + \xi_2\eta_2 + \ldots + \xi_n\eta_n.$$

Finalmente, los espacios complejos no degenerados de métrica simétrica hermitiana $(x, y) = \overline{(y, x)}$ se llaman *seudounitarios*. Del teorema 2 del p. 21.3 se deduce que *en todo espacio seudounitario de dimensión n existe una base en la que el producto escalar es de la forma*

$$(x, y) = \xi_1\bar{\eta}_1 + \ldots + \xi_s\bar{\eta}_s - \xi_{s+1}\bar{\eta}_{s+1} - \ldots - \xi_n\bar{\eta}_n.$$

El número $\sigma = s - (n - s)$ se llama *signatura* del espacio seudounitario. Con la dimensión n determina, obviamente, el espacio seudounitario, salvo un isomorfismo.

En cuanto a la clasificación de espacios bilineales métricos complejos no degenerados cualesquiera, ésta se obtiene en base al teorema 4 del p. 23.3. Para caracterizar el espacio es necesario en este caso escribir el conjunto de divisores elementales sujeto a las condiciones indicadas al final del p. 23.3.

24.3. Funciones bilineales en espacios bilineales métricos. En el punto anterior hemos logrado geometrizar la teoría de una función bilineal, definida sobre un espacio lineal $\mathfrak{L}$, gracias a que hemos considerado los valores de la función bilineal como los productos escalares de los vectores, definiendo así una métrica especial en $\mathfrak{L}$. Análogamente, para geometrizar la teoría de pares de funciones bilineales, definidas sobre un espacio lineal $\mathfrak{L}$, una de ellas se toma como la función métrica y la otra se considera como una función bilineal definida sobre un espacio bilineal métrico. Está claro que el par $\varphi(x, y)$, $\varphi_1(x, y)$ y el par $\varphi(x, y)$, $\varphi_2(x, y)$ serán ahora semejantes respecto a los automorfismos de un espacio *lineal* $\mathfrak{L}$ cuando,

y sólo cuando, la función $\varphi_1(x, y)$ puede ser reducida a la función $\varphi_2(x, y)$ mediante un automorfismo del espacio *bilineal métrico* $\mathfrak{L}$ de función métrica principal $\varphi(x, y)$.

Es más, en los espacios bilineales métricos no degenerados es posible relacionar biyectivamente con toda función bilineal una aplicación lineal del espacio. Puesto que esta relación no depende de cómo se escoja la base, resulta así que el estudio de las funciones bilineales equivale al estudio de las aplicaciones lineales de espacios bilineales métricos. Desde el punto de vista de la teoría de espacios lineales, esto significa que el estudio de pares de funciones bilineales, de las cuales una es no degenerada, equivale al estudio de pares formados por una función bilineal no degenerada y por una aplicación lineal.

Sea, pues, $\mathfrak{L}$ un espacio bilineal métrico no degenerado hermitiano o corriente.

TEOREMA 4. *Toda función lineal $f(x)$ definida sobre $\mathfrak{L}$ puede ser representada univocamente en la forma (x, a).*

Efectivamente, la condición $f(x) = (x, a)$ equivale al sistema de relaciones

$$(a_j, a) = f(a_j) \quad (j = 1, \ldots, n), \tag{12}$$

donde $a_1, \ldots, a_n$ es una base de $\mathfrak{L}$. Tomando $a = \xi_1 a_1 + \ldots + \xi_n a_n$ y considerando (12) como un sistema de ecuaciones respecto a $\xi_1, \ldots, \xi_n$, vemos que éste es un sistema de n ecuaciones lineales con n incógnitas de determinante diferente de cero, ya que el último es el determinante de la matriz de la forma principal en la base escogida. Por consiguiente, las ecuaciones (12) tienen una solución única.

Igual que esto ha sido hecho en el p. 18.2 para los espacios unitarios, el teorema 4 puede ser empleado para introducir en $\mathfrak{L}$ el concepto de la aplicación conjugada.

Consideremos una aplicación lineal cualquiera $\mathcal{A}$ de este espacio. La expresión $(x\mathcal{A}, a)$ representa para todo vector dado a una función lineal de x. Según el teorema 4, en el espacio $\mathfrak{L}$ existe un vector determinado b tal que

$$(x\mathcal{A}, a) = (x, b) \tag{13}$$

para todo x. Indiquemos por $\mathcal{A}^*$ la aplicación que transforma a en b. Entonces $b = a\mathcal{A}^*$ y la igualdad (13) puede ser representada en la forma

$$(x\mathcal{A}, a) = (x, a\mathcal{A}^*). \tag{14}$$

La aplicación $\mathcal{A}^*$ se llama *conjugada a la derecha* de $\mathcal{A}$. La propiedad (14) caracteriza plenamente la aplicación conjugada a la derecha. Efectivamente, si para una aplicación $\mathcal{B}$ tenemos $(x\mathcal{A}, a) = (x, a\mathcal{B})$ cualesquiera que sean a y x, comparando esta igualdad con (14)

encontramos la relación $(x, a\mathcal{A}^* - a\mathcal{B}) = 0$ de la cual resulta que $a\mathcal{A}^* - a\mathcal{B}$ es un vector isótropo del espacio $\mathfrak{L}$. Puesto que $\mathfrak{L}$ no contiene vectores isótropos no nulos, tenemos $a\mathcal{A}^* - a\mathcal{B} = o$ y $\mathcal{A}^* = \mathcal{B}$.

Repitiendo los razonamientos del p. 18.2 es fácil demostrar que $\mathcal{A}^*$ es lineal y que son válidas las fórmulas corrientes

$$(\mathcal{A} + \mathcal{B})^* = \mathcal{A}^* + \mathcal{B}^*, \tag{15}$$

$$(\mathcal{A}\mathcal{B})^* = \mathcal{B}^*\mathcal{A}^*, \tag{16}$$

$$(\alpha\mathcal{A}^*) = \alpha\mathcal{A}^* \tag{17}$$

para los espacios bilineales métricos corrientes y las fórmulas (15), (16) y

$$(\alpha\mathcal{A})^* = \bar{\alpha}\mathcal{A}^* \tag{18}$$

para los espacios bilineales métricos hermitianos.

Resolvamos ahora el problema sobre la relación que existe entre las matrices de la aplicación $\mathcal{A}$ y de su conjugada a la derecha $\mathcal{A}^*$. Tomemos para ello una base en $\mathfrak{L}$ e indiquemos las matrices de las aplicaciones $\mathcal{A}$ y $\mathcal{A}^*$ por A y B, respectivamente. Sea $\mathfrak{L}$ un espacio corriente. Basándonos en las fórmulas (6) del p. 24.1, obtenemos

$$(x\mathcal{A}, a) = [x\mathcal{A}]\, G\, [a]' = [x]\, AG\, [a]',$$
$$(x, a\mathcal{A}^*) = [x]\, G\, [a\mathcal{A}^*]' = [x]\, GB'\, [a]',$$

donde G es la matriz de Gram del espacio $\mathfrak{L}$ en la base escogida. De aquí

$$AG = GB' \quad \text{y} \quad B' = G^{-1} AG. \tag{19}$$

En el caso de espacios bilineales métricos hermitianos las fórmulas (19) se sustituyen por las relaciones

$$AG = G\overline{B}' \quad \text{y} \quad \overline{B}' = G^{-1} AG. \tag{20}$$

Hemos visto que en los espacios unitarios existen sistemas ortonormales de coordenadas. En estos sistemas la matriz de Gram se convierte en la matriz unidad y las relaciones (20) nos ofrecen el resultado conocido (p. 18.2): $\overline{B}' = A$.

Hemos definido hasta el momento sólo la aplicación conjugada a la derecha. Es obvio que de la misma forma se puede introducir también la aplicación conjugada a la izquierda. Es decir, sea $\mathcal{A}$ una aplicación lineal de $\mathfrak{L}$. Repitiendo los razonamientos realizados anteriormente, veremos que en $\mathfrak{L}$ existe una aplicación única $\mathcal{C}$ que cumple la igualdad

$$(x, a\mathcal{A}) = (x\mathcal{C}, a)$$

cualesquiera que sean x y a.

La aplicación $\mathcal{C}$ es lineal. Convedremos en llamarla *conjugada a la izquierda* de $\mathcal{A}$ y en indicarla por ${}^*\mathcal{A}$.

Si la métrica de $\mathfrak{L}$ *es simétrica o antisimétrica, las aplicaciones conjugadas a la derecha y a la izquierda de cualquier aplicación lineal coinciden, ya que*

$$(x\mathcal{A}^*, y) = \pm (y, x\mathcal{A}^*) = \pm (y\mathcal{A}, x) = (x, y\mathcal{A}) = (x^*\mathcal{A}, y).$$

Pasando a la consideración de las funciones bilineales sobre $\mathfrak{L}$, observemos ante todo que el valor de la expresión $(x\mathcal{A}, y)$, donde $\mathcal{A}$ es una aplicación lineal, representa una función bilineal y, respectivamente, una función bilineal hermitiana en $\mathfrak{L}$.

Si las aplicaciones lineales $\mathcal{A}$ *y* $\mathcal{B}$ *son diferentes, las funciones bilineales correspondientes* $(x\mathcal{A}, y)$ *y* $(x\mathcal{B}, y)$ *también son diferentes.*

En efecto, lo contrario significaría que $(x\mathcal{A}, y) = (x\mathcal{B}, y)$ para todos los x e y de $\mathfrak{L}$. De aquí resultaría $(x\mathcal{A} - x\mathcal{B}, y) = 0$ para todos los valores de y, es decir, $x\mathcal{A} - x\mathcal{B}$ sería un vector isótropo a la izquierda. Puesto que $\mathfrak{L}$ no contiene vectores isótropos no nulos, tendríamos $x\mathcal{A} = x\mathcal{B}$, es decir, $\mathcal{A} = \mathcal{B}$.

Probemos ahora que *toda función bilineal* $f(x, y)$ *definida sobre* $\mathfrak{L}$ *puede ser representada en la forma* $(x\mathcal{A}, y)$, *donde* $\mathcal{A}$ *es una aplicación lineal del espacio* $\mathfrak{L}$. En esta proposición debe comprenderse por $f(x, y)$ una función corriente, si $\mathfrak{L}$ es un espacio bilineal métrico corriente, y una función hermitiana, si $\mathfrak{L}$ es un espacio bilineal métrico hermitiano.

Efectivamente, para todo valor dado de y, $f(x, y)$ es una función lineal de x. En virtud del teorema 4, esto significa que para todo y existe un vector z, determinado unívocamente, tal que la relación $f(x, y) = (x, z)$ es válida para cualesquiera valores de x. Indiquemos por $\mathcal{B}$ la aplicación que transforma y en z; entonces

$$f(x, y) = (x, y\mathcal{B}). \tag{21}$$

En el caso de un espacio bilineal métrico hermitiano tenemos

$$f(x, \alpha y_1 + \beta y_2) = \bar{\alpha} f(x, y_1) + \bar{\beta} f(x, y_2) = \bar{\alpha}(x, y_1\mathcal{B}) + \bar{\beta}(x, y_2\mathcal{B}) = (x, \alpha(y_1\mathcal{B}) + \beta(y_2\mathcal{B})).$$

Por otro lado, debido a (21),

$$f(x, \alpha y_1 + \beta y_2) = (x, (\alpha y_1 + \beta y_2)\mathcal{B}).$$

De aquí tenemos

$$(\alpha y_1 + \beta y_2)\mathcal{B} = \alpha(y_1\mathcal{B}) + \beta(y_2\mathcal{B}),$$

es decir, $\mathcal{B}$ es una aplicación lineal. Lo mismo tiene lugar en el caso de los espacios bilineales métricos corrientes. Indicando por $\mathcal{A}$ la aplicación conjugada a la izquierda de $\mathcal{B}$, podemos representar (21) en la forma $f(x, y) = (x\mathcal{A}, y)$. Hemos obtenido, pues, el teorema siguiente:

TEOREMA 8. *Si* $\mathfrak{L}$ *es un espacio bilineal métrico no degenerado corriente o hermitiano, la expresión* $(x\mathcal{A}, y)$, *donde* $\mathcal{A}$ *es una aplicación*

lineal del espacio $\mathfrak{L}$, *representa una función bilineal corriente o hermitiana, respectivamente, sobre* $\mathfrak{L}$. *Viceversa, para toda función bilineal* $f(x, y)$ *corriente o, respectivamente, hermitiana definida sobre* $\mathfrak{L}$ *existe una aplicación lineal* $\mathcal{A}$ *de este espacio, determinada unívocamente, tal que* $f(x, y) = (x\mathcal{A}, y)$ *para cualesquiera* x *e* y *de* $\mathfrak{L}$.

El teorema 5 establece una correspondencia biyectiva entre las funciones bilineales y las aplicaciones lineales de los espacios bilineales métricos no degenerados que permite reducir el estudio de las funciones bilineales al estudio de las aplicaciones lineales de estos espacios. Desde el punto de vista de la teoría de pares, el teorema 5 significa que el estudio de los pares de funciones bilineales, al menos una de las cuales es regular, puede ser reducido al estudio de los pares mixtos compuestos de una función bilineal regular y de una aplicación lineal.

Veamos cómo están relacionadas las matrices de una función bilineal $f(x, y)$ y de su correspondiente aplicación lineal $\mathcal{A}$. Tomemos en $\mathfrak{L}$ una base cualquiera y sean F y A las matrices de la función $f(x, y)$ y de la aplicación $\mathcal{A}$, respectivamente. Tenemos, según la fórmula (6) del p. 24.1,

$$f(x, y) = [x]\, F\, [y]' \quad \text{y} \quad (x\mathcal{A}, y) = [x\mathcal{A}]\, G\, [y]' = [x]\, AG\, [y]',$$

de donde

$$F = AG, \tag{22}$$

donde G es la matriz de Gram del espacio $\mathfrak{L}$ en la base escogida. Es fácil comprobar que esta fórmula tiene lugar también en el caso de los espacios bilineales métricos hermitianos.

Una aplicación lineal $\mathcal{A}$ de un espacio bilineal métrico $\mathfrak{L}$ se llama *simétrica* si

$$(x\mathcal{A}, y) = (x, y\mathcal{A}) \tag{23}$$

cualesquiera que sean x e y de $\mathfrak{L}$. Una aplicación $\mathcal{A}$ se llama *antisimétrica* si

$$(x\mathcal{A}, y) = -(x, y\mathcal{A}). \tag{24}$$

Comparando (23) y (24) con las fórmulas que definen las aplicaciones conjugadas a la derecha y a la izquierda, obtenemos en lugar de (23) y (24) las relaciones equivalentes

$$\mathcal{A} = \mathcal{A}^* = {}^*\mathcal{A}, \tag{25}$$

$$\mathcal{A} = -\mathcal{A}^* = -{}^*\mathcal{A}. \tag{26}$$

Sea $\mathfrak{L}$ un espacio de métrica *simétrica* y sea $\mathcal{A}$ una aplicación lineal *simétrica* del espacio $\mathfrak{L}$. La función bilineal correspondiente a la aplicación $\mathcal{A}$ es de la forma

$$f(x, y) = (x\mathcal{A}, y).$$

Si el espacio $\mathfrak{L}$ es corriente se tiene

$$f(x, y) = (x\mathcal{A}, y) = (y, x\mathcal{A}) = (y\mathcal{A}, x) = f(y, x).$$

Si el espacio $\mathfrak{L}$ es hermitiano, tenemos respectivamente

$$f(x, y) = (x\mathcal{A}, y) = \overline{(y, x\mathcal{A})} = \overline{(y\mathcal{A}, x)} = \overline{f(y, x)}.$$

Por consiguiente, $f(x, y)$ es en ambos casos una función simétrica. Los mismos razonamientos muestran que la simetría de $f(x, y)$ implica la simetría de la correspondiente aplicación lineal $\mathcal{A}$. Análogamente, la antisimetría de la aplicación $\mathcal{A}$ equivale a la antisimetría de la correspondiente función $f(x, y)$.

Si la métrica del espacio $\mathfrak{L}$ es antisimétrica, la relación entre la simetría y la antisimetría de las aplicaciones lineales y de las funciones bilineales es la inversa: a las aplicaciones simétricas les corresponden funciones antisimétricas y a las aplicaciones antisimétricas les corresponden las funciones simétricas. Efectivamente, si $\mathcal{A}$ es una aplicación simétrica de un espacio bilineal métrico corriente $\mathfrak{L}$ provisto de una métrica antisimétrica, tenemos

$$f(x, y) = (x\mathcal{A}, y) = -(y, x\mathcal{A}) = -(y\mathcal{A}, x) = -f(y, x).$$

Análogamente se demuestran las demás afirmaciones. Hemos llegado así al teorema siguiente:

TEOREMA 6. *Si $\mathfrak{L}$ es un espacio no degenerado corriente o hermitiano provisto de una métrica simétrica, a las funciones bilineales simétricas les corresponden aplicaciones lineales simétricas del espacio $\mathfrak{L}$ y a las funciones antisimétricas les corresponden aplicaciones antisimétricas. Si la métrica del espacio $\mathfrak{L}$ es antisimétrica, al contrario, a las funciones antisimétricas les corresponden aplicaciones simétricas y a las funciones simétricas les corresponden aplicaciones antisimétricas.*

Volviendo a la teoría de pares, vemos del teorema 6 que el estudio de los pares de funciones bilineales simétricas o antisimétricas equivale al estudio de las aplicaciones lineales simétricas o antisimétricas en los espacios de métrica simétrica o antisimétrica.

Ejemplos y problemas

1. La matriz de Gram de un espacio métrico bilineal $\mathfrak{L}$ en una base a_1, a_2, a_3 y a_4 es igual a

$$\begin{bmatrix} 1 & -3 & -7 & -2 \\ 2 & 1 & 7 & 3 \\ 3 & -2 & 0 & 1 \\ 4 & -1 & 5 & 3 \end{bmatrix}.$$

Demuéstrese que el subespacio isótropo a la izquierda de $\mathfrak{L}$ tiene la base $a_1 + a_2 - a_3$ y $5a_2 + 6a_3 - 7a_4$ y que el subespacio isótropo a la derecha tiene la base $2a_1 + 3a_2 - a_3$ y $a_1 + a_2 - a_4$.

2. Para que un espacio bilineal métrico se descomponga en una suma directa de subespacios ortogonales a ambos lados es necesario y suficiente que su matriz de Gram se descomponga en una base.

3. Si la matriz de Gram de un espacio bilineal métrico no degenerado $\mathfrak{L}$ es G, la aplicación lineal de matriz $G'G^{-1}$ del espacio $\mathfrak{L}$ es un automorfismo del mismo.

4. Demuéstrese que cualquiera que sea un subespacio no degenerado $\mathfrak{A}$ de un espacio bilineal métrico $\mathfrak{L}$, las dimensiones de $\mathfrak{A}^{\perp}$ y de ${}^{\perp}\mathfrak{A}$ son iguales a la diferencia de las dimensiones de $\mathfrak{L}$ y de $\mathfrak{A}$. Demuéstrese, en particular, que ${}^{\perp}(\mathfrak{A}^{\perp})=({}^{\perp}\mathfrak{A})^{\perp}=\mathfrak{A}$.

5. Todo espacio provisto de una métrica simétrica o antisimétrica es una suma directa de unos subespacios isótropo y no degenerado.

6. Si en un espacio bilineal métrico $\mathfrak{L}$ son equivalentes los conceptos de ortogonalidad a la derecha y a la izquierda, $\mathfrak{L}$ es o bien un espacio de métrica simétrica o bien un espacio de métrica antisimétrica.

7. Las funciones ψ que cumplen la identidad

$$\beta\psi(\alpha x+y)+\alpha\psi(x-\beta y)=(1+\alpha\beta)(\alpha\psi(x)+\beta\psi(y)),$$

y sólo estas funciones, son cuadráticas sobre un espacio lineal.

8. Sea I una matriz cuadrada fija. Una matriz A se llama *I-ortogonal*, si $AIA'=I$, se llama *I-simétrica*, si $AI=IA'$, y se llama *I-antisimétrica*, si $AI=-IA'$. Demuéstrese que si en una base de un espacio bilineal métrico la matriz de Gram coincide con I, en esta base las aplicaciones isométricas tienen matrices I-ortogonales y las aplicaciones simétricas y antisimétricas tienen matrices I-simétricas e I-antisimétricas, respectivamente.

Capítulo VII

Aplicaciones lineales de espacios bilineales métricos

En el capítulo presente será considerada la clasificación de los principales tipos de aplicaciones lineales (simétricas, antisimétricas e isométricas) de espacios provistos de métrica bilineal. La relación que existe entre esta clasificación y la clasificación de los pares de formas bilineales ha sido explicada en el § 24 del cap. VI y se supone que el lector que pase al estudio de este capítulo está familiarizado con los resultados de aquel parágrafo.

En este capítulo, al igual que en el anterior, se supone que todos los espacios que aquí aparecen son espacios sobre un cuerpo conmutativo (pero no sobre un cuerpo cualquiera).

§ 25. Tipos principales de aplicaciones lineales

25.1. Automorfismos. Según el p. 24.2, una aplicación lineal regular $\mathcal{U}$ de un espacio bilineal métrico $\mathfrak{L}$ se llama *automorfismo* del espacio $\mathfrak{L}$, si $\mathcal{U}$ no altera la magnitud del producto escalar, es decir, si

$$(x\mathcal{U}, y\mathcal{U}) = (x, y) \tag{1}$$

para todos los x e y de $\mathfrak{L}$. Los automorfismos del espacio $\mathfrak{L}$ también se llaman a veces aplicaciones *isométricas* del mismo. Empleando los conceptos de las aplicaciones conjugadas a la derecha y a la izquierda, podemos representar la relación (1) en la forma

$$(x, y) = (x\mathcal{U}, y\mathcal{U}) = (x, y\mathcal{U}\mathcal{U}^*) = (x\mathcal{U}^*\mathcal{U}, y),$$

de donde se tiene

$$\mathcal{U}\mathcal{U}^* = \mathcal{U}^*\mathcal{U} = \mathcal{E}. \tag{2}$$

Está claro que, recíprocamente, la relación (2) implica la relación (1). Por consiguiente, *para que una aplicación lineal $\mathcal{U}$ de un espacio bilineal métrico no degenerado sea un automorfismo es necesario y*

suficiente que ambas aplicaciones conjugadas de $\mathcal{U}$ coincidan con la inversa de $\mathcal{U}$.

Tomemos en $\mathfrak{L}$ una base e indiquemos por U la matriz de la aplicación $\mathcal{U}$. Si el espacio $\mathfrak{L}$ es corriente, la relación (1) se convierte en

$$[x]\,UGU'\,[y]' = [x]\,G\,[y]',$$

donde G es la matriz de Gram. De aquí resulta

$$UGU' = G. \tag{3}$$

Si el espacio $\mathfrak{L}$ es hermitiano, de (1) se tiene

$$[x]\,UG\overline{U}'\,[\overline{y}]' = [x]\,G\,[\overline{y}]',$$

es decir,

$$UG\overline{U}' = G. \tag{4}$$

Las igualdades (3) y (4) representan las condiciones que deben cumplir *las matrices de las aplicaciones isométricas* de los espacios bilineales métricos no degenerados corrientes y hermitianos, respectivamente.

Dos aplicaciones lineales $\mathcal{A}_1$ y $\mathcal{A}_2$ de un espacio bilineal métrico $\mathfrak{L}$ se llaman *isomorfas*, si existe una aplicación isomorfa $\mathcal{U}$ del espacio $\mathfrak{L}$ sobre sí mismo que transforma $\mathcal{A}_1$ en $\mathcal{A}_2$. Tenemos, de acuerdo con el p. 10.2,

$$\mathcal{A}_2 = \mathcal{U}^{-1}\mathcal{A}_1\mathcal{U}. \tag{5}$$

La aplicación $\mathcal{U}$ es un automorfismo del espacio $\mathfrak{L}$ y, por ello, la relación (5) significa que *el isomorfismo de las aplicaciones lineales de espacios bilineales métricos equivale a la semejanza de las mismas respecto a las aplicaciones isométricas.* De aquí se deduce, en virtud de los resultados del p. 15.3, que las aplicaciones lineales isomorfas tienen factores invariantes iguales. En el caso general, este criterio no es suficiente para el isomorfismo de unas aplicaciones. Sin embargo, la situación es diferente, si se consideran aplicaciones simétricas, antisimétricas o isométricas.

TEOREMA 1. *Sea $\mathfrak{L}$ un espacio no degenerado corriente sobre el cuerpo de todos los números complejos provisto de una métrica simétrica o antisimétrica. Entonces para el isomorfismo de unas aplicaciones simétricas, antisimétricas o isométricas del espacio $\mathfrak{L}$ es necesario y suficiente que los factores invariantes de estas aplicaciones coincidan.*

La necesidad ha sido demostrada anteriormente y, por esto, consideraremos sólo la suficiencia. Sean $\mathcal{A}_1$ y $\mathcal{A}_2$ las aplicaciones lineales dadas. Por hipótesis, los factores invariantes de $\mathcal{A}_1$ y de $\mathcal{A}_2$ coinciden y, por consiguiente,

$$\mathcal{A}_2 = \mathcal{T}^{-1}\mathcal{A}_1\mathcal{T}, \tag{6}$$

donde $\mathscr{T}$ es una aplicación lineal regular del espacio $\mathfrak{L}$. Pasando en ambos miembros a las aplicaciones conjugadas, obtenemos [1]

$$\mathscr{A}_2^* = \mathscr{T}^* \mathscr{A}_1^* \mathscr{T}^{*-1}. \tag{7}$$

Si $\mathscr{A}_1$ y $\mathscr{A}_2$ son simétricas o antisimétricas, se tiene $\mathscr{A}_1^* = \pm \mathscr{A}_1$ y $\mathscr{A}_2^* = \pm \mathscr{A}_2$ y (7) se convierte en

$$\mathscr{A}_2 = \mathscr{T}^* \mathscr{A}_1 \mathscr{T}^{*-1} \tag{8}$$

En cambio, si $\mathscr{A}_1$ y $\mathscr{A}_2$ son isométricas, tenemos $\mathscr{A}_1^* = \mathscr{A}_1^{-1}$ y $\mathscr{A}_2^* = \mathscr{A}_2^{-1}$. Introduciendo estos valores en (7) y elevando el resultado a la potencia -1, obtenemos de nuevo (8). Debido a (6) y (8) tenemos $\mathscr{T}^{-1} \mathscr{A}_1 \mathscr{T} = \mathscr{T}^* \mathscr{A}_1 \mathscr{T}^{*-1}$, de donde

$$\mathscr{A}_1 \cdot \mathscr{T} \mathscr{T}^* = \mathscr{T} \mathscr{T}^* \cdot \mathscr{A}_1. \tag{9}$$

De (9) se desprende directamente que

$$\mathscr{A}_1 \cdot (\mathscr{T} \mathscr{T}^*)^k = (\mathscr{T} \mathscr{T}^*)^k \cdot \mathscr{A}_1 \quad (k = 1, 2, \ldots)$$

y, en general,

$$\mathscr{A}_1 \cdot f(\mathscr{T} \mathscr{T}^*) = f(\mathscr{T} \mathscr{T}^*) \cdot \mathscr{A}_1,$$

donde $f(\lambda)$ es un polinomio arbitrario (compárese con el p. 23.2). Según el teorema sobre la extracción de la raíz cuadrada (p. 16.3), el polinomio $f(\lambda)$ se puede escoger de modo que

$$f(\mathscr{T} \mathscr{T}^*) \cdot f(\mathscr{T} \mathscr{T}^*) = \mathscr{T} \mathscr{T}^*.$$

Tomando

$$\mathscr{D} = f(\mathscr{T} \mathscr{T}^*) \text{ y } \mathscr{U} = \mathscr{D}^{-1} \mathscr{T},$$

obtenemos

$$\mathscr{D}^* = [f(\mathscr{T} \mathscr{T}^*)]^* = f(\mathscr{T} \mathscr{T}^*) = \mathscr{D},$$

$$\mathscr{U}^* = \mathscr{T}^* \mathscr{D}^{*-1} = \mathscr{T}^* \mathscr{D}^{-1},$$

$$\mathscr{U} \mathscr{U}^* = \mathscr{D}^{-1} \mathscr{T} \mathscr{T}^* \mathscr{D}^{-1} = \mathscr{D}^{-1} f(\mathscr{T} \mathscr{T}^*)^2 \mathscr{D}^{-1} = \mathscr{D}^{-1} \mathscr{D}^2 \mathscr{D}^{-1} = \mathscr{E}.$$

Por consiguiente, $\mathscr{U}$ es una aplicación isométrica. Al mismo tiempo de $\mathscr{A}_1 \mathscr{D} = \mathscr{D} \mathscr{A}_1$ se deduce que

$$\mathscr{U}^{-1} \mathscr{A}_1 \mathscr{U} = \mathscr{T}^{-1} \mathscr{D} \mathscr{A}_1 \mathscr{D}^{-1} \mathscr{T} = \mathscr{T}^{-1} \mathscr{A}_1 \mathscr{D} \mathscr{D}^{-1} \mathscr{T} = \mathscr{T}^{-1} \mathscr{A}_1 \mathscr{T} = \mathscr{A}_2$$

que es lo que se quería demostrar.

El teorema 1 muestra que en los espacios euclídeos complejos, así como en los espacios complejos simpliciales, para clasificar, salvo un isomorfismo, las aplicaciones simétricas, antisimétricas e isométricas es suficiente saber qué divisores elementales pueden contener estas aplicaciones.

---

[1] Como la métrica de $\mathfrak{L}$ es simétrica o antisimétrica, las aplicaciones conjugadas a la derecha y a la izquierda coinciden.

TEOREMA 2. *Si una aplicación isométrica $\mathcal{U}$ de un espacio bilineal métrico no degenerado corriente $\mathfrak{L}$ contiene k veces el divisor elemental $(\lambda-\alpha)^m$, la aplicación $\mathcal{U}$ contiene también k veces el divisor elemental $(\lambda-\alpha^{-1})^m$.*

La afirmación del teorema no tiene contenido para $\alpha = \pm 1$. Por esto aceptaremos que $\alpha \neq \pm 1$. Tomemos en $\mathfrak{L}$ una base e indiquemos por U la matriz de la aplicación $\mathcal{U}$. En virtud de (3), tenemos $UGU' = G$, donde G es la matriz de Gram. De aquí resulta

$$U' = G^{-1}U^{-1}G. \tag{10}$$

Los divisores elementales de la matriz U' coinciden con los divisores elementales de la matriz U. La fórmula (10) muestra que, a su vez, los divisores elementales de la matriz U' coinciden con los divisores elementales de la matriz U^{-1} (compárese con el p. 15.3). Es decir, los divisores elementales de la matriz U deben ser los mismos que los de la matriz U^{-1}. Pero, según el teorema sobre los divisores elementales de las funciones (p. 16.4), los divisores elementales de la matriz U^{-1} se obtienen de los divisores elementales $(\lambda-\alpha)^m$ de la matriz U sustituyendo α por α^{-1}. Por consiguiente, si en todo divisor elemental de tipo $(\lambda-\alpha)^m$ de la matriz U sustituimos α por α^{-1}, obtendremos de nuevo un divisor elemental de la matriz U que es lo que se quería demostrar.

En el caso de espacios bilineales métricos hermitianos la relación (10) se convierte en

$$\overline{U}' = G^{-1}U^{-1}G.$$

Con arreglo a esto también cambia la afirmación sobre los divisores elementales: *si $(\lambda-\alpha)^m$ es un divisor elemental de multiplicidad k de una aplicación isométrica de un espacio bilineal métrico no degenerado hermitiano, la expresión $(\lambda-\bar{\alpha}^{-1})^m$ también será un divisor elemental de multiplicidad k de esta aplicación.* La demostración es la misma.

TEOREMA 3. *Sean a y b dos vectores radicales de una aplicación isométrica $\mathcal{U}$ de un espacio bilineal métrico corriente $\mathfrak{L}$. Si los valores propios α y β, a los que corresponden estos vectores, no son recíprocamente inversos, es decir, si $\alpha\beta \neq 1$, los vectores a y b son ortogonales. Análogamente, si $\mathfrak{L}$ es un espacio bilineal métrico hermitiano y $\alpha\bar{\beta} \neq 1$, los vectores a y b son también ortogonales.*

La demostración es la misma tanto para los espacios corrientes como para los hermitianos. Por esto consideraremos sólo los espacios corrientes. Por hipótesis,

$$a(\alpha\mathcal{E} - \mathcal{U})^s = o \quad \text{y} \quad b(\beta\mathcal{E} - \mathcal{U})^t = o \quad (\alpha\beta \neq 1), \tag{11}$$

donde s y t son unos números enteros no negativos. Debemos demostrar que de las relaciones (11) se desprende la igualdad $(a, b) = 0$. Realizaremos la demostración por inducción según los valores de

la suma $s+t$. Si $s+t=0$, se tiene $s=t=0$ y de las relaciones (11) resulta $a=b=o$, de donde $(a, b)=0$.

Supongamos ahora que tenemos dado un valor de la suma $s+t$ y que para todos los valores menores que esta suma la afirmación ha sido ya demostrada. Tomemos

$$a(\alpha\mathscr{E}-\mathcal{U})=a_1 \text{ y } b(\beta\mathscr{E}-\mathcal{U})=b_1.$$

Puesto que

$$a_1(\alpha\mathscr{E}-\mathcal{U})^{s-1}=a(\alpha\mathscr{E}-\mathcal{U})^s=o,$$
$$b_1(\beta\mathscr{E}-\mathcal{U})^{t-1}=b(\beta\mathscr{E}-\mathcal{U})^t=o,$$

tenemos por inducción

$$(a, b_1)=(a_1, b)=(a_1, b_1)=0.$$

Pero la igualdad $(a, b_1)=0$ implica que $(a, b(\beta\mathscr{E}-\mathcal{U}))=0$, es decir,

$$(a, b\mathcal{U})=\beta(a, b). \tag{12}$$

Análogamente, de la igualdad $(a_1, b)=0$ se deduce:

$$(a\mathcal{U}, b)=\alpha(a, b). \tag{13}$$

Finalmente, de la igualdad $(a_1, b_1)=0$ resulta:

$$0=(a(\alpha\mathscr{E}-\mathcal{U}),\ b(\beta\mathscr{E}-\mathcal{U}))=$$
$$=\alpha\beta(a, b)-\alpha(a, b\mathcal{U})-\beta(a\mathcal{U}, b)+(a\mathcal{U}, b\mathcal{U}),$$

de donde, debido a (12) y (13), obtenemos

$$0=-\alpha\beta(a, b)+(a, b). \tag{14}$$

Puesto que $\alpha\beta\neq 1$, la igualdad (14) ofrece la relación requerida $(a, b)=0$.

TEOREMA 4. *Si $\mathcal{A}$ es una aplicación isométrica, simétrica o antisimétrica del espacio $\mathfrak{L}$ y el subespacio $\mathfrak{A}$ es invariante respecto a $\mathcal{A}$, los subespacios ortogonales a $\mathfrak{A}$ a la derecha y a la izquierda son también invariantes respecto a $\mathcal{A}$.*

Sea $\mathcal{A}$ una aplicación isométrica del espacio $\mathfrak{L}$. El subespacio ortogonal a la derecha $\mathfrak{A}^{\perp}$ está formado por los vectores b que cumplen la relación $(a, b)=0$ cualquiera que sea a de $\mathfrak{A}$. Puesto que $\mathcal{A}$ es una aplicación regular, el subespacio $\mathfrak{A}$ es invariante también respecto a $\mathcal{A}^{-1}$. Por consiguiente, el vector $a\mathcal{A}^{-1}$ pertenece a $\mathfrak{A}$, de donde tenemos $(a\mathcal{A}^{-1}, b)=0$. Pero

$$(a\mathcal{A}^{-1}, b)=(a\mathcal{A}^{-1}\mathcal{A}, b\mathcal{A})=(a, b\mathcal{A});$$

es decir, $(a, b\mathcal{A})=0$ para todo a de $\mathfrak{A}$. Por consiguiente, el vector $b\mathcal{A}$ figura en $\mathfrak{A}^{\perp}$ y el subespacio $\mathfrak{A}^{\perp}$ es invariante respecto a $\mathcal{A}$. También sencillas son las demostraciones de las demás afirmaciones.

TEOREMA 5. *Supongamos que un espacio bilineal métrico corriente o hermitiano $\mathfrak{L}$ se descompone en una suma directa de sus subespacios*

$\mathfrak{L}_1, \mathfrak{L}_2, \ldots, \mathfrak{L}_s$ *ortogonales a ambos lados. Si todos estos subespacios son invariantes respecto a una aplicación lineal* $\mathcal{A}$ *del espacio* $\mathfrak{L}$ *y si la aplicación* $\mathcal{A}$ *es isométrica o, respectivamente, simétrica o antisimétrica sobre cada uno de los subespacios* $\mathfrak{L}_1, \ldots, \mathfrak{L}_s$, *la aplicación* $\mathcal{A}$ *será del mismo tipo sobre el espacio* $\mathfrak{L}$.

Este teorema explica cómo pueden emplearse las descomposiciones directas para el estudio de las propiedades de las aplicaciones lineales de los tipos señalados. Su demostración es evidente y queda a cargo del lector.

25.2. Aplicaciones simétricas y antisimétricas. Recordemos que una aplicación $\mathcal{A}$ de un espacio bilineal métrico corriente o hermitiano se llama *simétrica* si

$$(x\mathcal{A}, y) = (x, y\mathcal{A})$$

para cualesquiera x e y de $\mathfrak{L}$.

TEOREMA 6. *Los vectores radicales* a *y* b *correspondientes a diferentes valores propios* ρ *y* σ *de una aplicación simétrica* $\mathcal{A}$ *de un espacio bilineal métrico corriente o hermitiano son ortogonales.*

Efectivamente, tenemos, por hipótesis,

$$a(\rho\mathcal{E} - \mathcal{A})^s = o \quad \text{y} \quad b(\sigma\mathcal{E} - \mathcal{A})^t = o, \tag{15}$$

donde s y t son unos números enteros no negativos. Debemos demostrar que de (15) se desprende la ortogonalidad de los vectores a y b. Realizaremos la demostración por inducción según los valores de la suma $s+t$. Para $s+t=1$ o bien s o bien t es igual a cero y, por consiguiente, o bien $a=o$ o bien $b=o$, de donde $(a, b)=0$.

Supongamos ahora que tenemos dado un valor de la suma $s+t$ y que para todos los valores menores que esta suma la afirmación ha sido ya demostrada. Sea

$$a_1 = a(\rho\mathcal{E} - \mathcal{A}) \quad \text{y} \quad b_1 = b(\sigma\mathcal{E} - \mathcal{A}).$$

Tenemos

$$a_1(\rho\mathcal{E} - \mathcal{A})^{s-1} = a(\rho\mathcal{E} - \mathcal{A})^s = o,$$
$$b_1(\sigma\mathcal{E} - \mathcal{A})^{t-1} = b(\sigma\mathcal{E} - \mathcal{A})^t = o.$$

Por esto, de acuerdo con la hipótesis de inducción, resulta

$$(a, b_1) = (a_1, b) = 0,$$

es decir,

$$\sigma(a, b) = (a, b\mathcal{A}) \quad \text{y} \quad \rho(a, b) = (a\mathcal{A}, b).$$

Puesto que la aplicación $\mathcal{A}$ es simétrica, los segundos miembros de estas igualdades coinciden y, por consiguiente, $(\sigma-\rho)(a, b)=0$, de donde se tiene $(a, b)=0$ que es lo que se quería demostrar.

Una aplicación $\mathcal{A}$ de un espacio bilineal métrico corriente que cumple la relación

$$(x\mathcal{A}, y) = -(x, y\mathcal{A})$$

para todos los x e y de $\mathfrak{L}$ se llama, de acuerdo con el p. 24.3, *antisimétrica*.

TEOREMA 7. *Sean a y b unos vectores radicales de una aplicación antisimétrica de un espacio bilineal métrico corriente correspondientes a los valores propios* ρ *y* σ. *Si* $\rho + \sigma \neq 0$, *se tiene* $(a, b) = 0$.

La demostración coincide casi textualmente con las demostraciones análogas de los teoremas 3 y 6 y puede ser aquí omitida.

Las aplicaciones isométricas de un espacio bilineal métrico no degenerado cualquiera $\mathfrak{L}$ están estrechamente ligadas a las aplicaciones antisimétricas del mismo. Tiene lugar concretamente el siguiente teorema que, por la forma en la que se enuncia, coincide plenamente con el teorema del p. 20.3 sobre la aplicación de Cayley.

TEOREMA 8. *Sea* $\mathfrak{L}$ *un espacio bilineal métrico corriente o hermitiano. Si* $\mathcal{A}$ *es una aplicación antisimétrica del espacio* $\mathfrak{L}$ *que no tiene valores propios iguales a* -1, *la aplicación*

$$\mathcal{U} = (\mathcal{E} - \mathcal{A})(\mathcal{E} + \mathcal{A})^{-1} \tag{16}$$

es una aplicación isométrica del espacio $\mathfrak{L}$ *que tampoco tiene valores propios iguales a* -1 *y, además,* $\mathcal{A}$ *se expresa mediante* $\mathcal{U}$ *por la fórmula*

$$\mathcal{A} = (\mathcal{E} - \mathcal{U})(\mathcal{E} + \mathcal{U})^{-1}. \tag{17}$$

Recíprocamente, si $\mathcal{U}$ *es una aplicación isométrica del espacio* $\mathfrak{L}$ *que no tiene valores propios iguales a* -1, *la aplicación* $\mathcal{A}$ *calculada mediante la fórmula* (17) *es antisimétrica y tampoco tiene valores propios iguales a* -1 *y* $\mathcal{U}$ *se expresa en términos de* $\mathcal{A}$ *mediante la fórmula* (16).

Igual que en el caso de espacios unitarios, las fórmulas (16) y (17) llevan el nombre de *aplicaciones de Cayley*. La demostración de las mismas coincide textualmente con la demostración de estas fórmulas realizada en el p. 20.3 para los espacios unitarios. Por esto omitimos aquí la demostración. Observemos que unas fórmulas análogas

$$\mathcal{U} = -(\mathcal{E} - \mathcal{A})(\mathcal{E} + \mathcal{A})^{-1}, \tag{18}$$

$$\mathcal{A} = (\mathcal{E} + \mathcal{U})(\mathcal{E} - \mathcal{U})^{-1} \tag{19}$$

ofrecen una correspondencia entre las aplicaciones antisimétricas del espacio $\mathfrak{L}$ que no tienen valores propios iguales a -1 y las aplicaciones isométricas del espacio $\mathfrak{L}$ que no tienen valores propios iguales a $+1$.

Las aplicaciones de Cayley podrían reducir totalmente el estudio de las aplicaciones isométricas al estudio de las antisimétricas, si no figurasen los valores propios excepcionales ± 1. La existencia de estos valores hace necesario el estudio independiente más detallado de las propiedades de las aplicaciones isométricas.

Para terminar demostremos que tiene lugar el teorema siguiente:

TEOREMA 9. *Si $(\lambda-\alpha)^m$ figura k veces en el sistema de divisores elementales de una aplicación antisimétrica $\mathcal{A}$ de un espacio bilineal métrico no degenerado corriente $\mathfrak{L}$, la expresión $(\lambda+\alpha)^m$ también figura k veces en el sistema de los divisores elementales de la aplicación $\mathcal{A}$.*

Efectivamente, la matriz B de la aplicación conjugada $\mathcal{A}^*$ satisface la relación (19) del p. 24.3:

$$B' = G^{-1}AG,$$

donde G es la matriz de Gram y A es la matriz de la aplicación $\mathcal{A}$. De aquí se ve que los divisores elementales de la matriz B coinciden con los divisores elementales de la matriz A. Sin embargo, de la condición de antisimetría se deduce que $B = -A$; por consiguiente, cambiando el signo de α en cada uno de los divisores elementales $(\lambda-\alpha)^m$ de la aplicación $\mathcal{A}$, obtendremos de nuevo unos divisores elementales de la aplicación $\mathcal{A}$ que es lo que se quería demostrar.

Ejemplos y problemas

1. Sea $\mathcal{U}$ una aplicación de un espacio bilineal métrico no degenerado $\mathfrak{L}$ sobre sí mismo que conserva el producto escalar $(a\mathcal{U}, b\mathcal{U}) = (a, b)$. Demuéstrese que $\mathcal{U}$ es una aplicación lineal y, por consiguiente, isométrica del espacio $\mathfrak{L}$.
2. Los determinantes de las aplicaciones isométricas de los espacios bilineales métricos no degenerados corrientes son iguales a ± 1 y los determinantes de las aplicaciones isométricas de los espacios bilineales métricos no degenerados hermitianos son de módulo igual a la unidad.
3. En todo espacio bilineal métrico no degenerado $\mathfrak{L}$ existe una aplicación lineal $\mathcal{S}$ que satisface la condición $(x, y) = (y\mathcal{S}, x)$ si $\mathfrak{L}$ es un espacio corriente, y la condición $(x, y) = \overline{(y\mathcal{S}, x)}$, si $\mathfrak{L}$ es un espacio hermitiano. Demuéstrese que la aplicación $\mathcal{S}$ es isométrica y que su matriz es igual a $G'G^{-1}$ o $\overline{G}'G^{-1}$ según sea $\mathfrak{L}$ un espacio corriente o hermitiano (G es una matriz de Gram del espacio $\mathfrak{L}$).
4. Si en un espacio bilineal métrico no degenerado corriente $\mathfrak{L}$ coinciden las aplicaciones conjugadas a la derecha y a la izquierda de cualquier aplicación lineal, la métrica del espacio $\mathfrak{L}$ es simétrica o antisimétrica.

§ 26. Espacios euclídeos complejos

En el parágrafo presente y en los dos que le siguen examinaremos más detalladamente las formas elementales a las que pueden ser reducidas las matrices de las aplicaciones simétricas, antisimétricas e isométricas de los espacios euclídeos, simpliciales y seudounitarios sobre el cuerpo de los números complejos; estos últimos han sido clasificados de un modo completo al final del p. 24.2. Notemos que se tratará de determinar las formas elementales de las matrices de las aplicaciones en unos sistemas especiales de coorde-

nadas con la matriz de Gram bien definida, ya que sólo en esta condición se conocen, salvo un isomorfismo, las aplicaciones. En el caso de los espacios euclídeos unitarios y reales este problema ha sido resuelto en el cap. V. Para el sistema especial de coordenadas ha sido escogido en aquella ocasión un sistema ortonormal de coordenadas en el que la matriz de Gram es la matriz unidad. En el caso de espacios de tipo más complejo es preferible tomar para los sistemas principales o normales de coordenadas sistemas de coordenadas con una matriz de Gram de estructura más compleja. Comenzaremos por el estudio de un espacio euclídeo complejo.

26.1. Aplicaciones simétricas. Como ya hemos señalado, se llama espacio euclídeo complejo un espacio no degenerado sobre el cuerpo de los números complejos provisto de una métrica simétrica corriente. La matriz de Gram de un espacio euclídeo complejo es regular y simétrica. Recíprocamente, todo espacio bilineal métrico complejo con la matriz de Gram regular simétrica es un espacio euclídeo complejo. Pero todos los espacios euclídeos complejos de una dimensión n dada son isomorfos y por esto en todo espacio de esta índole existe un sistema de coordenadas con una matriz de Gram simétrica regular cualquiera dada de antemano. En particular, en todo espacio euclídeo complejo $\mathfrak{L}$ existe un sistema de coordenadas $a_1, a_2, \ldots, a_n$ en el que la matriz de Gram es de la forma

$$G = \begin{bmatrix} 0 & \ldots & 0 & 1 \\ 0 & \ldots & 1 & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 1 & \ldots & 0 & 0 \end{bmatrix}. \tag{1}$$

Convendremos en llamar *normal* todo sistema de coordenadas que tenga esta matriz. Los vectores de un sistema normal de coordenadas satisfacen las relaciones

$$\begin{gathered} (a_j, a_{n+1-j}) = 1, \quad (a_j, a_k) = 0 \\ (j + k \neq n + 1; \quad j, k = 1, \ldots, n), \end{gathered} \tag{2}$$

que también son suficientes para que un sistema sea normal.

La conveniencia de las bases normales se determina por la siguiente propiedad de las mismas: *toda aplicación lineal, cuya matriz en una base normal es una célula de Jordan, es simétrica.*

Efectivamente, por lo visto en el p. 24.3, una aplicación $\mathcal{A}$ es simétrica si es simétrica la matriz AG, donde G es la matriz de Gram de la base y A es la matriz de la aplicación $\mathcal{A}$ en esta base. Pero realizando directamente los cálculos se puede ver que al multiplicar una célula de Jordan de orden n por la matriz G se obtiene una matriz simétrica.

Esta observación permite demostrar inmediatamente el siguiente teorema principal:

TEOREMA 1. *Sea dado un sistema arbitrario de expresiones de tipo* $(\lambda-\rho_1)^{m_1}, \ldots, (\lambda-\rho_s)^{m_s}$. *En todo espacio euclídeo complejo de dimensión* $n=m_1+\ldots+m_s$ *existe entonces necesariamente una aplicación simétrica* $\mathcal{A}$ *para la cual las expresiones señaladas constituyen el sistema completo de sus divisores elementales.*

En efecto, pongamos

$$G_j=\begin{bmatrix} 0 & \ldots & 0 & 1 \\ 0 & \ldots & 1 & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 1 & \ldots & 0 & 0 \end{bmatrix} \quad \text{y} \quad A_j=\begin{bmatrix} \rho_j & 1 & 0 & \ldots \\ & \rho_j & 1 & \ldots \\ & & \ldots & \ldots \\ & & & \rho_j \end{bmatrix}, \tag{3}$$

donde los órdenes de las matrices G_j y A_j son iguales a m_j $(j=1, \ldots, s)$. Sea

$$G=G_1\dotplus\ldots\dotplus G_s \quad \text{y} \quad A=A_1\dotplus\ldots\dotplus A_s. \tag{4}$$

En un espacio euclídeo complejo $\mathfrak{L}$ de dimensión n existe una base de matriz de Gram igual a G. Indiquemos por $\mathcal{A}$ la aplicación lineal de matriz A en la base señalada. Debido a la observación hecha anteriormente, la matriz AG es simétrica y con ella es simétrica también la aplicación $\mathcal{A}$. Al mismo tiempo de (3) y de (4) se ve que $\mathcal{A}$ tiene el conjunto requerido de divisores elementales.

En virtud del teorema 1 del p.25.1, el teorema demostrado resuelve completamente el problema sobre la clasificación, salvo un isomorfismo, de todas las aplicaciones simétricas de un espacio euclídeo complejo. En particular, de él se desprende el teorema siguiente:

TEOREMA 2. *Sea* $\mathcal{A}$ *una aplicación simétrica de un espacio euclídeo complejo* $\mathfrak{L}$. *Entonces,* $\mathfrak{L}$ *se puede descomponer en una suma directa de subespacios recíprocamente ortogonales invariantes respecto de* $\mathcal{A}$ *y tales que en cada uno de ellos existe una base normal en la que la matriz de la aplicación inducida por la aplicación* $\mathcal{A}$ *será una célula de Jordan.*

Para la demostración indiquemos por $(\lambda-\rho_1)^{m_1}, \ldots, (\lambda-\rho_s)^{m_s}$ el conjunto de los divisores elementales de la aplicación $\mathcal{A}$. Tomemos en $\mathfrak{L}$ una base $a_1, \ldots, a_n$ tal que su matriz de Gram sea igual a la matriz G de (4), y sea $\mathcal{B}$ la aplicación lineal de matriz A de (4) en la base $a_1, \ldots, a_n$. Las aplicaciones $\mathcal{A}$ y $\mathcal{B}$ son simétricas y semejantes. En virtud del mencionado teorema 1 del p. 25.1, de aquí se deduce que existe una aplicación isométrica $\mathcal{U}$ tal que $\mathcal{A}=\mathcal{U}\mathcal{B}\mathcal{U}^{-1}$. Entonces, la matriz de la aplicación $\mathcal{A}$ en la base $a_1\mathcal{U}, \ldots, a_n\mathcal{U}$ coincidirá con la matriz de la aplicación $\mathcal{B}$ en la base $a_1, \ldots, a_n$, es decir, coincidirá con la matriz A. Al mismo tiempo, la matriz de Gram de la base $a_1\mathcal{U}, \ldots, a_n\mathcal{U}$ es G y para la aplicación $\mathcal{B}$ las afirmaciones del teorema 2 son evidentes.

En una base ortonormal la matriz de una aplicación simétrica es simétrica. Por ello, del teorema 1 se deduce que existen matrices complejas simétricas de cualquier conjunto de divisores elementales.

26.2 Aplicaciones antisimétricas. Sea $\mathcal{A}$ una aplicación antisimétrica de un espacio euclídeo complejo $\mathfrak{L}$. Descompongamos $\mathfrak{L}$ en la suma directa de los subespacios radicales

$$\mathfrak{L} = \mathfrak{L}_{\rho_1} \dotplus \mathfrak{L}_{\rho_2} \dotplus \ldots \dotplus \mathfrak{L}_{\rho_s}.$$

Agrupando los sumandos que corresponden a pares opuestos de valores propios, obtenemos una nueva descomposición de $\mathfrak{L}$:

$$\mathfrak{L} = \mathfrak{M}_0 \dotplus \mathfrak{M}_1 \dotplus \ldots \dotplus \mathfrak{M}_t,$$

donde $\mathfrak{M}_0$ es el subespacio radical correspondiente a la raíz cero y $\mathfrak{M}_j (j = 1, \ldots, t)$ son sumas de tipo $\mathfrak{L}_\rho \dotplus \mathfrak{L}_{-\rho}$. El teorema 7 del p.25.2 muestra que los subespacios $\mathfrak{M}_0, \ldots, \mathfrak{M}_t$ son recíprocamente ortogonales y, además, son invariantes respecto de $\mathcal{A}$. Por esto el estudio del comportamiento de $\mathcal{A}$ sobre $\mathfrak{L}$ se reduce al estudio del comportamiento de esta aplicación sobre los espacios $\mathfrak{M}_0, \mathfrak{M}_1, \ldots, \mathfrak{M}_t$ por separado. Consideremos más detalladamente el subespacio $\mathfrak{M}_0$. Supongamos que $\mathfrak{M}_0$ es diferente de cero e indiquemos por $\mathcal{A}_0$ la aplicación inducida en $\mathfrak{M}_0$ por la aplicación $\mathcal{A}$. Puesto que todos los valores propios de la aplicación $\mathcal{A}_0$ son iguales a cero, se tiene $\mathcal{A}_0^p = \mathcal{O}$, donde p es la dimensión de $\mathfrak{M}_0$. Indiquemos por m la menor potencia de la aplicación $\mathcal{A}_0$ que se convierte en cero: $\mathcal{A}_0^m = \mathcal{O}$ y $\mathcal{A}_0^{m-1} \neq \mathcal{O}$. Consideremos dos casos: el de m par y el de m impar. Sea m par; entonces $m-1$ es impar y la aplicación $\mathcal{A}_0^{m-1}$ es antisimétrica. La función bilineal $(x\mathcal{A}_0^{m-1}, y)$ es también antisimétrica. Puesto que esta función es diferente de cero, en $\mathfrak{M}_0$ existe un par de vectores a y b para el cual

$$(a\mathcal{A}_0^{m-1}, b) = 1.$$

De la antisimetría de la aplicación $\mathcal{A}_0$ se desprende que

$$(a\mathcal{A}_0^k, b\mathcal{A}_0^l) = -(a\mathcal{A}_0^{k-1}, b\mathcal{A}_0^{l+1}) = \ldots (-1)^k (a, b\mathcal{A}_0^{l+k}).$$

En particular, se tiene $(a, b\mathcal{A}_0^{m-1}) = (-1)^{m-1} (a\mathcal{A}_0^{m-1}, b) = -1$, es decir $b\mathcal{A}_0^{m-1} \neq o$. Pongamos

$$a_1 = a, \quad a_2 = a_1\mathcal{A}_0, \quad \ldots, \quad a_m = a_{m-1}\mathcal{A}_0,$$
$$b_1 = b, \quad b_2 = b_1\mathcal{A}_0, \quad \ldots, \quad b_m = b_{m-1}\mathcal{A}_0.$$

Las relaciones

$$(a_k, a_{m+1-k}) = (-1)^{k-1} (a_1, a_1\mathcal{A}_0^{m-1}) = 0, \quad (b_k, b_{m+1-k}) = 0,$$
$$(a_k, b_{m+1-k}) = (-1)^{m-k}, \quad (a_k, b_j) = 0$$
$$(k + j > m + 1; \quad k, j = 1, \ldots, m)$$

muestran que la matriz de Gram de esta base es de la forma

$$\begin{bmatrix} * & \dots & 0 & * & \dots & -1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & 1 & \dots & 0 \\ * & \dots & -1 & * & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & \dots & 0 & 0 & \dots & 0 \end{bmatrix},$$

donde en cada una de las cuatro células todos los elementos que figuran debajo de la diagonal secundaria son iguales a cero. El determinante de tal matriz es igual a ± 1 y, por consiguiente, el subespacio $\mathfrak{N}_1$ tendido sobre los vectores $a_1, \dots, a_m, b_1, \dots, b_m$ es no degenerado. Debido al teorema 2 del p. 24.2, se tiene

$$\mathfrak{M}_0 = \mathfrak{N}_1 \dot{+} \mathfrak{N}_1^{\perp},$$

siendo $\mathfrak{N}_1^{\perp}$ de nuevo un subespacio invariante respecto de $\mathcal{A}_0$.

Consideremos el caso en que m es impar. La aplicación $\mathcal{A}_0^{m-1}$ es ahora simétrica y, por consiguiente, es simétrica su correspondiente función bilineal $(x\mathcal{A}_0^{m-1}, y)$. Puesto que esta función no es igual idénticamente a cero, existe un vector a tal que $(a\mathcal{A}_0^{m-1}, a) \neq 0$. Pongamos

$$a_1 = a, \quad a_2 = a_1\mathcal{A}_0, \quad \dots, \quad a_m = a_{m-1}\mathcal{A}_0.$$

El subespacio $\mathfrak{N}_1$ tendido sobre los vectores $a_1, \dots, a_m$ es invariante respecto de $\mathcal{A}_0$. Su matriz de Gram en el sistema de coordenadas $a_1, \dots, a_m$ es de la forma

$$G = \begin{bmatrix} (a_1, a_1) & \dots & (a_1, a_{m-1}) & (a_1, a_m) \\ (a_2, a_1) & \dots & (a_2, a_{m-1}) & 0 \\ \dots & \dots & \dots & \dots \\ (a_m, a_1) & \dots & 0 & 0 \end{bmatrix}.$$

Efectivamente,

$$(a_j, a_k) = (a\mathcal{A}_0^{j-1}, a\mathcal{A}_0^{k-1}) = (-1)^{k-1}(a\mathcal{A}_0^{j+k-2}, a) = 0$$

para $j + k > m + 1$. Además

$$(a_1, a_m) = -(a_2, a_{m-1}) = \dots = (a_m, a_1) = \alpha \neq 0.$$

De aquí se ve que la matriz G es regular; por consiguiente, el subespacio $\mathfrak{N}_1$ es no degenerado y tenemos de nuevo

$$\mathfrak{M}_0 = \mathfrak{N}_1 \dot{+} \mathfrak{N}_1^{\perp},$$

donde $\mathfrak{N}_1^{\perp}$ es invariante respecto de $\mathcal{A}_0$.

Es decir, hemos despejado de $\mathfrak{M}_0$ un subespacio $\mathfrak{N}_1$ que en el primer caso representa una suma de dos subespacios de dimensión par y en el segundo caso es un subespacio de dimensión impar.

Aplicando este mismo procedimiento al subespacio complementario $\mathfrak{M}_1^{\perp}$ podremos descomponerlo de nuevo en una suma directa de subespacios invariantes, etc. El espacio $\mathfrak{M}_0$ resultará descompuesto de esta forma en una suma directa de subespacios invariantes, cada uno de los cuales será o bien una suma directa de dos subespacios de dimensión par o bien un subespacio de dimensión impar. En esta descomposición a todo subespacio corresponde un divisor elemental de la aplicación $\mathcal{A}_0$ de forma λ^p, donde p es la dimensión del subespacio. Por consiguiente, *toda aplicación antisimétrica de un espacio euclídeo complejo contiene un número par de veces a todo divisor elemental de tipo* λ^{2s}. Este resultado, así como el teorema 7 del p. 25.2, imponen ciertas condiciones al sistema de divisores elementales de una aplicación antisimétrica. Demostremos que estas condiciones son suficientes para la existencia de una aplicación antisimétrica.

TEOREMA 3. *Las aplicaciones antisimétricas de los espacios euclídeos complejos contienen los divisores elementales correspondientes a los valores propios no nulos en forma de pares* $(\lambda-\alpha)^m$, $(\lambda+\alpha)^m$, *los divisores elementales de tipo* λ^p *también en forma de pares* λ^p, λ^p *para p par y los divisores elementales de tipo* λ^p *para p impar en combinaciones arbitrarias. Recíprocamente, todo conjunto formado por un número finito de expresiones de tipo* $(\lambda-\alpha_i)^{m_i}$ *que cumple estas condiciones, es un sistema de divisores elementales de una aplicación antisimétrica de un espacio euclídeo complejo de dimensión adecuada.*

La primera parte de esta proposición ha sido ya demostrada. Por ello, para obtener la demostración completa del teorema debemos construir para cualquier sistema de expresiones, que cumple las condiciones del teorema, la correspondiente aplicación antisimétrica. Por analogía con el punto anterior esto se puede lograr del modo siguiente. A todo par de expresiones $(\lambda-\alpha_i)^{m_i}$, $(\lambda+\alpha_i)^{m_i}$ y entre ellos a los pares con $\alpha_i=0$ ponemos en correspondencia las matrices

$$G_i=\begin{bmatrix} 0 & D_i \\ D_i & 0 \end{bmatrix} \quad \text{y} \quad A_i=\begin{bmatrix} B_i & 0 \\ 0 & -B_i \end{bmatrix},$$

donde D_i es una matriz simétrica *normal* de tipo (1) del p. 26.1 y B_i es la célula de Jordan de orden m_i y con el valor propio a_i. Si entre las expresiones dadas figuran expresiones λ^{2s_i-1} que no forman pares, les ponemos en correspondencia las matrices

$$G_i=\begin{bmatrix} 0 & \dots & 0 & 1 \\ 0 & \dots & 1 & 0 \\ \dots & \dots & \dots & \dots \\ 1 & \dots & 0 & 0 \end{bmatrix} \quad \text{y} \quad A_i=\begin{bmatrix} 0 & 1 & 0 & 0 & \dots & & \\ & 0 & -1 & 0 & \dots & & \\ & & & 0 & 1 & \dots & \\ \dots & & & & & & \\ & & & & & 0 & -1 \\ & & & & & & 0 \end{bmatrix}$$

de orden $2s_i - 1$. Sea

$$G = G_1 + \ldots + G_s \quad \text{y} \quad A = A_1 + \ldots + A_s. \tag{5}$$

Indiquemos por $\mathfrak{L}$ un espacio complejo de matriz de Gram igual a G.

Puesto que G es una matriz simétrica regular, el espacio $\mathfrak{L}$ es un espacio complejo euclídeo. Consideremos la aplicación lineal $\mathcal{A}$ del espacio $\mathfrak{L}$, cuya matriz es A. La función bilineal correspondiente a la aplicación $\mathcal{A}$ tiene la matriz AG con la particularidad de que

$$AG = A_1G_1 + \ldots + A_sG_s.$$

Los cálculos directos muestran que las células A_iG_i, y por ello también la matriz AG, son antisimétricas. Por esto la función bilineal y la aplicación $\mathcal{A}$ son también antisimétricas. Se ve, de la forma de la matriz A, que los divisores elementales de la aplicación $\mathcal{A}$ tienen los valores requeridos.

Según el teorema 1 del p. 25.1, todas las aplicaciones antisimétricas de un espacio euclídeo complejo que tengan sistemas iguales de divisores elementales son isomorfas. Luego, del teorema 3 se deduce que para toda aplicación antisimétrica $\mathcal{A}$ de un espacio euclídeo complejo $\mathfrak{L}$ existe un sistema de coordenadas en el que la matriz de Gram G y la matriz A de la aplicación son de la forma (5).

26.3. Aplicaciones ortogonales complejas. Las aplicaciones isométricas de un espacio euclídeo complejo $\mathfrak{L}$ suelen llamarse aplicaciones *ortogonales complejas*. Si en $\mathfrak{L}$ se ha tomado un sistema de coordenadas de matriz de Gram G, las matrices U de las aplicaciones ortogonales, y sólo estas matrices, satisfacen la relación

$$UGU' = G. \tag{6}$$

En particular, si el sistema de coordenadas es ortonormal, se tiene $G = E$ y (6) se convierte en

$$UU' = E.$$

En otras palabras, *en un sistema ortonormal de coordenadas las matrices de las aplicaciones ortogonales son ortogonales.*

Las aplicaciones ortogonales con los divisores elementales iguales son, según el teorema 1 del p. 25.1, isomorfas. Por esto para la clasificación de las aplicaciones ortogonales es suficiente indicar qué sistemas de expresiones de tipo $(\lambda - \alpha_i)^{m_i}$ pueden servir como sistemas de divisores elementales de las aplicaciones ortogonales. El esquema de la solución de este problema es el siguiente: conocemos los divisores elementales de las aplicaciones antisimétricas; las aplicaciones ortogonales se pueden expresar en términos de las antisimétricas mediante las fórmulas de Cayley (p. 25.2); por consiguiente,

empleando el teorema sobre los divisores elementales de las funciones matriciales (p. 16.4), podemos encontrar también los divisores elementales de las aplicaciones ortogonales. Sin embargo, al realizar este esquema debe tenerse en cuenta la existencia de los valores excepcionales en las fórmulas de Cayley, debido a lo cual la solución adquiere la forma siguiente, algo más voluminosa.

Sea $\mathcal{U}$ una aplicación ortogonal de un espacio euclídeo complejo $\mathfrak{L}$. Representando $\mathfrak{L}$ en la forma de la suma directa de los subespacios radicales de la aplicación $\mathcal{U}$ y agrupando los subespacios correspondientes a valores propios recíprocamente inversos, obtendremos la descomposición

$$\mathfrak{L} = \mathfrak{M}_{-1} \dotplus \mathfrak{M}_1 \dotplus \mathfrak{M}_2 \dotplus \ldots \dotplus \mathfrak{M}_l,$$

donde todos los $\mathfrak{M}_j$ son recíprocamente ortogonales y, además, $\mathfrak{M}_{-1}$ y $\mathfrak{M}_1$ son los subespacios radicales de la aplicación $\mathcal{U}$ correspondientes a los valores propios -1 y $+1$. La aplicación $\mathcal{U}$ induce en cada uno de los subespacios $\mathfrak{M}_j$ una aplicación ortogonal $\mathcal{U}_j$ y los divisores elementales de la aplicación $\mathcal{U}$ se descomponen en los sistemas de divisores elementales de estas aplicaciones inducidas. Consideremos la aplicación $\mathcal{U}_1$. Todos sus valores propios son iguales a $+1$. La fórmula de Cayley

$$\mathcal{A}_1 = (\mathcal{E} - \mathcal{U}_1)(\mathcal{E} + \mathcal{U}_1)^{-1}$$

transforma $\mathcal{U}_1$ en una aplicación antisimétrica $\mathcal{A}_1$. Representemos esta fórmula en la forma $\mathcal{A}_1 = f(\mathcal{U}_1)$, donde $f(\lambda) = (1-\lambda)(1+\lambda)^{-1}$. Puesto que la derivada de $f(\lambda)$ no se anula para $\lambda = 1$, resulta (p. 16.4) que introduciendo en todo divisor elemental $(\lambda-\alpha)^m$ de la aplicación $\mathcal{U}_1$ en lugar de α el número $(1-\alpha)(1+\alpha)^{-1}$, obtendremos los divisores elementales de la aplicación $\mathcal{A}_1$. Pero los divisores elementales de la aplicación $\mathcal{U}_1$ son de la forma $(\lambda-1)^s$; por consiguiente, los divisores elementales de la aplicación $\mathcal{A}_1$ son de tipo λ^s. Si s es un número par, las aplicaciones antisimétricas contienen todo divisor elemental de tipo λ^s un número par de veces y por esto la aplicación $\mathcal{U}_1$ contiene todo divisor elemental de tipo $(\lambda-1)^s$ también un número par de veces, si s es par.

Aplicando la fórmula de Cayley

$$\mathcal{A} = (\mathcal{E} + \mathcal{U})(\mathcal{E} - \mathcal{U})^{-1}$$

a la aplicación $\mathcal{U}_{-1}$, obtendremos, de forma análoga, que la aplicación $\mathcal{U}$ contiene todo divisor elemental de tipo $(\lambda+1)^s$ también un número par de veces, si s es par.

TEOREMA 4. *Los divisores elementales de las aplicaciones ortogonales complejas correspondientes a los valores propios diferentes de* ± 1 *aparecen en pares de tipo* $(\lambda-\alpha)^m$, $(\lambda-\alpha^{-1})^m$; *los divisores elementales de tipo* $(\lambda \pm 1)^{2s}$ *aparecen un número par de veces y los divisores elementales de tipo* $(\lambda \pm 1)^{2s+1}$ *aparecen en combinaciones arbi-*

trarias. Recíprocamente, todo sistema de expresiones de tipo $(\lambda-\alpha_i)^{m_i}$ $(\alpha_i \neq 0)$ *que cumple estas condiciones es un sistema de divisores elementales de una aplicación ortogonal compleja.*

La primera parte del teorema ha sido ya demostrada. Supongamos, por ello, que tenemos un sistema de divisores elementales de tipo $(\lambda-\alpha_i)^{m_i}(\alpha_i \neq 0)$ que satisface las condiciones del teorema. Separamos en este sistema los divisores elementales de tipo $(\lambda+1)^p$ y para cada uno de los divisores elementales $(\lambda-\alpha_i)^{m_i}$ que quedan construimos la expresión $(\lambda-\beta_i)^{m_i}$, donde $\beta_i=(1-\alpha_i)(1+\alpha_i)^{-1}$. El sistema de las expresiones $(\lambda-\beta_i)^{m_i}$ cumplirá las condiciones del teorema 3 y, por consiguiente, será un sistema de divisores elementales de una aplicación antisimétrica $\mathcal{A}_0$ de un espacio euclídeo complejo $\mathfrak{L}_0$. Puesto que hemos tomado solamente valores de α_i diferentes de -1, todos los β_i serán también diferentes de -1. Empleando para $\mathcal{A}_0$ la aplicación de Cayley, obtenemos una aplicación ortogonal $\mathcal{U}_0$ con los divisores elementales $(\lambda-\alpha_i)^{m_i}(\alpha_i \neq -1)$. Análogamente, para todo divisor elemental de tipo $(\lambda+1)^{m_i}$ tomamos la expresión λ^{m_i} y buscamos una aplicación antisimétrica $\mathcal{A}_1$ de un espacio $\mathfrak{L}_1$, cuyos divisores elementales sean λ^{m_i}. Entonces la aplicación

$$\mathcal{U}_1=-(\mathcal{E}-\mathcal{A}_1)(\mathcal{E}+\mathcal{A}_1)^{-1}$$

será una aplicación ortogonal del espacio $\mathfrak{L}_1$, cuyos divisores elementales son $(\lambda+1)^{m_i}$. Tomemos en $\mathfrak{L}_0$ y en $\mathfrak{L}_1$ unos sistemas ortonormales de coordenadas e indiquemos por U_0 y U_1 las matrices de las aplicaciones $\mathcal{U}_0$ y $\mathcal{U}_1$. Estas matrices serán ortogonales y por ello la suma directa $U_0 \dot{+} U_1 = U$ de las mismas también será una matriz ortogonal.

Los divisores elementales de la matriz U tendrán, con arreglo a la construcción, los valores requeridos y el teorema queda demostrado.

Ejemplos y problemas

1. Constrúyase una matriz ortogonal de divisores elementales $(\lambda-1)^2$, $(\lambda-1)^2$, $\lambda-\frac{1}{2}$ y $\lambda-2$.

2. Demuéstrese que las matrices de tipo $\lambda A+B$, donde A es una matriz simétrica regular y B es una matriz antisimétrica, contienen un número par de veces todo divisor elemental de tipo λ^{2m}.

3. Empleando el teorema 2, redúzcase a la forma elemental el siguiente par de formas cuadráticas

$$2\xi_1^2+2\xi_2^2+2\xi_1\xi_3+2\xi_2\xi_4-2\xi_1\xi_3-2\xi_3\xi_4,$$
$$\xi_1^2+3\xi_2^2+2\xi_3^2+2\xi_1\xi_3-4\xi_2\xi_3+2\xi_3\xi_4-2\xi_3\xi_4.$$

Respuesta: $2(\xi_1\xi_2+\xi_3\xi_4)$ y $\xi_1^2+2\xi_3\xi_4+\xi_4^2$.

4. Constrúyase una matriz compleja antisimétrica de divisores elementales $(\lambda-3)^3$, $(\lambda+3)^3$, λ^2, λ^3 y λ^3.

5. Constrúyase una matriz compleja simétrica de divisores elementales $(\lambda-2)^3$, λ^2 y λ^3.

§ 27. Espacios simpliciales

27.1. Aplicaciones simétricas. Se llaman espacios simpliciales, según el p. 24.2, los espacios bilineales métricos provistos de una métrica antisimétrica no degenerada. La dimensión de un espacio simplicial es un número par. Puesto que todos los espacios simpliciales de una dimensión dada son recíprocamente isomorfos, en todo espacio simplicial de dimensión $n=2m$ existe una base, cuya matriz de Gram es igual a cualquier matriz antisimétrica regular de orden n dada de antemano. En particular, en todo espacio simplicial existen sistemas de coordenadas, en los cuales la matriz de Gram es de la forma

$$S=\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}+\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}+\cdots+\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \qquad (1)$$

Estos sistemas se llamarán *simpliciales*. Análogamente, en $\mathfrak{L}$ existen sistemas de coordenadas, en los cuales la matriz de Gram es de la forma

$$J=\begin{bmatrix} O & E \\ -E & O \end{bmatrix},$$

donde E es una matriz unidad y O es una matriz nula. Convendremos en llamar estos sistemas *normales*. En lo que sigue consideraremos solamente espacios simpliciales *complejos*.

Sea ahora $\mathcal{A}$ una aplicación simétrica de $\mathfrak{L}$. Descompongamos $\mathfrak{L}$ en la suma directa

$$\mathfrak{L}=\mathfrak{L}_{\rho_1}+\mathfrak{L}_{\rho_2}+\cdots+\mathfrak{L}_{\rho_s},$$

donde $\rho_1, \rho_2, \ldots, \rho_s$ son los diferentes valores propios de la aplicación $\mathcal{A}$ y $\mathfrak{L}_{\rho_1}, \ldots, \mathfrak{L}_{\rho_s}$ son los correspondientes subespacios radicales. Los subespacios $\mathfrak{L}_{\rho_i}$ son, según el p. 25.2, recíprocamente ortogonales. Debido a que el espacio $\mathfrak{L}$ es no degenerado, los subespacios $\mathfrak{L}_{\rho_1}, \ldots, \mathfrak{L}_{\rho_s}$ también son no degenerados y, por consiguiente, simpliciales. Tomemos uno de ellos, digamos $\mathfrak{L}_{\rho_i}$, e indiquémoslo por $\mathfrak{M}$. Sea $\mathcal{A}_0$ la aplicación inducida en $\mathfrak{M}$ por la aplicación $\mathcal{A}$. Pongamos, además, $\mathcal{A}_0-\rho_i\mathcal{E}=\mathcal{B}$. Puesto que $\mathcal{A}_0$ y $\mathcal{E}$ son aplicaciones simétricas, $\mathcal{B}$ es también una aplicación simétrica. Todos los valores propios de la aplicación $\mathcal{A}_0$ son iguales a ρ_i; luego, indicando por p la dimensión del espacio $\mathfrak{M}$, obtenemos

$$(\mathcal{A}_0-\rho_i\mathcal{E})^p=\mathcal{B}^p=\mathcal{O}.$$

Sea m el exponente menor para el cual $\mathcal{B}^m = \mathcal{O}$. La expresión $(x\mathcal{B}^{m-1}, y)$ es una función bilineal antisimétrica en x e y correspondiente a la aplicación $\mathcal{B}^{m-1}$. Puesto que $\mathcal{B}^{m-1} \neq \mathcal{O}$, la función $(x\mathcal{B}^{m-1}, y)$ no puede ser igual idénticamente a cero y, por consiguiente, en $\mathfrak{M}$ existe un par de vectores a y b tal que $(a\mathcal{B}^{m-1}, b) = 1$. Pongamos

$$a\mathcal{B}^j = a_{j+1} \text{ y } b\mathcal{B}^j = b_{j+1} \ (j = 0, 1, \ldots, m-1).$$

Debido a la simetría de $\mathcal{B}$, tenemos

$$(a_j, a_k) = (a\mathcal{B}^{j-1}, a\mathcal{B}^{k-1}) = (a\mathcal{B}^{j+k-2}, a) = (a, a\mathcal{B}^{j+k-2}).$$

Sin embargo, el espacio $\mathfrak{L}$ es antisimétrico y por esto

$$(a\mathcal{B}^{j+k-2}, a) = -(a, a\mathcal{B}^{j+k-2}).$$

Comparando esta fórmula con la anterior, vemos que

$$(a_j, a_k) = 0 \ (j, k = 1, 2, \ldots, m). \tag{2}$$

Análogamente se obtienen también las relaciones

$$(b_j, b_k) = 0 \ (j, k = 1, 2, \ldots, m). \tag{3}$$

Además, de las condiciones $(a\mathcal{B}^{m-1}, b) = 1$ y $\mathcal{B}^m = \mathcal{O}$ se deduce:

$$(a_{m-s}, b_{s+1}) = (a\mathcal{B}^{m-s-1}, b\mathcal{B}^s) = (a\mathcal{B}^{m-1}, b) = (a_m, b_1) = 1 \tag{4}$$

y

$$(a_j, b_k) = (a\mathcal{B}^{j-1}, b\mathcal{B}^{k-1}) = (a\mathcal{B}^{j+k-2}, b) = 0 \ (j+k > m+1). \tag{5}$$

Formando ahora la matriz de Gram de los vectores $a_1, \ldots, a_m, b_1, \ldots, b_m$, veremos, en virtud de las relaciones de (2) a (5), que tiene la forma triangular con ceros sobre la segunda diagonal. Puesto que los elementos que se hallan en la propia segunda diagonal son iguales a ± 1, el determinante de esta matriz es diferente de cero. Por consiguiente, el subespacio $\mathfrak{M}_1$ tendido sobre los vectores $a_1, \ldots, a_m, b_1, \ldots, b_m$ es no degenerado. Indiquemos por $\mathfrak{M}_{11}$ y $\mathfrak{M}_{12}$ los subespacios tendidos sobre los vectores $a_1, \ldots, a_m$ y $b_1, \ldots, b_m$, respectivamente. Debido a las relaciones (5), los subespacios $\mathfrak{M}_{11}$ y $\mathfrak{M}_{12}$ son de una misma dimensión y $\mathfrak{M}_1$ es la suma directa de los mismos. Puesto que $\mathfrak{M}_1$ es no degenerado, tenemos, en virtud del punto 24.2,

$$\mathfrak{M} = \mathfrak{M}_1 + \mathfrak{M}_1^{\perp}.$$

El subespacio $\mathfrak{M}_1$ es invariante respecto a $\mathcal{B}$ y por esto $\mathfrak{M}_1^{\perp}$ también es invariante respecto a $\mathcal{B}$. Procediendo ahora con $\mathfrak{M}_1^{\perp}$ igual que con $\mathfrak{M}$, podemos despejar en $\mathfrak{M}_1^{\perp}$ un subespacio $\mathfrak{M}_2$ que es una suma directa de dos subespacios $\mathfrak{M}_{21}$ y $\mathfrak{M}_{22}$, etc. Después de un número finito de pasos obtendremos la descomposición directa

$$\mathfrak{M} = \mathfrak{M}_1 + \mathfrak{M}_2 + \ldots + \mathfrak{M}_t,$$

donde cada uno de los subespacios $\mathfrak{M}_1, \ldots, \mathfrak{M}_t$ será, a su vez, una suma directa de dos subespacios de una misma dimensión. A todo subespacio de dimensión m que figura en esta descomposición le corresponde un divisor elemental $(\lambda-\rho_i)^m$ de la aplicación $\mathcal{A}$. Por consiguiente, llegamos a la conclusión de que en un espacio simplicial los divisores elementales de las aplicaciones simétricas aparecen en pares $(\lambda-\rho_i)^m$, $(\lambda-\rho_i)^m$.

Demostremos ahora que, recíprocamente, para todo sistema de expresiones $(\lambda-\rho_i)^{m_i}$, $(\lambda-\rho_i)^{m_i}$ existe una aplicación simétrica de un espacio simplicial $\mathfrak{L}$, cuyos divisores elementales son estas expresiones.

A todo par $(\lambda-\rho_i)^{m_i}$, $(\lambda-\rho_i)^{m_i}$ ponemos en correspondencia el par de matrices

$$G_i=\begin{bmatrix} O & E_i \\ -E_i & O \end{bmatrix} \quad \text{y} \quad A_i=\begin{bmatrix} B_i & O \\ O & B_i \end{bmatrix},$$

donde E_i es la matriz unidad de orden m_i y B_i es la célula de Jordan de orden m_i con el valor propio ρ_i y sea

$$G=G_1+\ldots+G_s \quad \text{y} \quad A=A_1+\ldots+A_s. \tag{6}$$

Consideremos un espacio bilineal métrico complejo $\mathfrak{L}$ de matriz de Gram G. Puesto que G es una matriz antisimétrica regular, el espacio $\mathfrak{L}$ es un espacio simplicial. Los cálculos directos muestran que

$$AG=A_1G_1+A_2G_2+\ldots+A_sG_s$$

es una matriz antisimétrica. Por esto la aplicación lineal $\mathcal{A}$ del espacio $\mathfrak{L}$, cuya matriz es A, es simétrica. Al mismo tiempo, los divisores elementales de la aplicación $\mathcal{A}$ tienen los valores requeridos. Es decir, hemos obtenido el teorema siguiente:

TEOREMA 1. *Todo divisor elemental de una aplicación simétrica de un espacio simplicial aparece un número par de veces. Recíprocamente, todo sistema de expresiones de tipo* $(\lambda-\rho_i)^{m_i}$ *que cumple esta condición es un sistema de divisores elementales de una aplicación simétrica de un espacio simplicial.*

Sabemos ya (p. 25.1) que las aplicaciones simétricas de los espacios simpliciales que tienen divisores elementales iguales son isomorfas. Por ello, el teorema 1 resuelve plenamente el problema de clasificación de las aplicaciones simétricas. En particular, muestra que para toda aplicación simétrica $\mathcal{A}$ de un espacio simplicial $\mathfrak{L}$ se puede escoger una base tal que la matriz de Gram G y la matriz A de la aplicación sean de la forma (6).

27.2. Aplicaciones antisimétricas. Una aplicación lineal $\mathcal{A}$ de un espacio bilineal métrico $\mathfrak{L}$ es antisimétrica, si $(x\mathcal{A}, y)=-(x, y\mathcal{A})$ para todos los x e y de $\mathfrak{L}$.

En virtud del teorema 3 del p. 26.2, los divisores elementales de una aplicación antisimétrica $\mathcal{A}$ de un espacio simplicial $\mathfrak{L}$, correspondientes a los valores propios no nulos, aparecen en forma de pares de tipo $(\lambda-\rho)^m$, $(\lambda+\rho)^m$. Probemos que los divisores elementales de la aplicación $\mathcal{A}$ correspondientes al valor propio nulo, es decir, los que son de forma λ^m, también aparecen en forma de pares, λ^m, λ^m, si m es impar. La demostración es completamente análoga a la demostración de la afirmación respectiva del p. 26.2 y sólo indicaremos el esquema de la misma.

Nos basamos en la representación de $\mathfrak{L}$ en la forma de la suma directa de los subespacios radicales de la aplicación $\mathcal{A}$. Agrupando en esta representación los subespacios correspondientes a los valores propios opuestos, obtendremos la descomposición

$$\mathfrak{L}=\mathfrak{M}_0\dotplus\mathfrak{M}_1\dotplus\ldots\dotplus\mathfrak{M}_s,$$

donde todos los subespacios $\mathfrak{M}_0, \ldots, \mathfrak{M}_s$ son invariantes y recíprocamente ortogonales, siendo $\mathfrak{M}_0$ el subespacio radical correspondiente a la raíz nula. Sea $\mathcal{A}_0$ la aplicación del subespacio $\mathfrak{M}_0$ inducida en éste por la aplicación $\mathcal{A}$. Todos los valores propios de la aplicación $\mathcal{A}_0$ son iguales a cero y, por ello, $\mathcal{A}_0^p=\mathcal{O}$, donde p es la dimensión de $\mathfrak{M}_0$. Indiquemos por m el menor exponente para el cual $\mathcal{A}_0^m=\mathcal{O}$. La aplicación $\mathcal{A}_0$ es antisimétrica y, por consiguiente, la aplicación $\mathcal{A}_0^{m-1}$ será, igual que en el p. 26.2, simétrica para m impar y antisimétrica para m par. Sin embargo, la correspondiente función bilineal $(x\mathcal{A}_0^{m-1}, y)$ será ahora simétrica para m par y antisimétrica para m impar, ya que la métrica de los espacios simpliciales es antisimétrica. Los razonamientos ulteriores del p. 26.2 se conservan con la única diferencia que se obtendrá ahora un par de espacios en el caso de m impar.

TEOREMA 2. *Los divisores elementales de las aplicaciones antisimétricas de los espacios simpliciales correspondientes a los valores propios no nulos aparecen en forma de pares* $(\lambda-\rho)^m$, $(\lambda+\rho)^m$; *los divisores elementales de tipo* λ^m *con m impar también aparecen en forma de pares* λ^m, λ^m *y los divisores elementales* λ^m *con m par pueden aparecer en combinaciones arbitrarias. Recíprocamente, todo sistema de expresiones de tipo* $(\lambda-\rho_i)^{m_i}$ *que posee estas propiedades es un sistema de divisores elementales de una aplicación antisimétrica de un espacio simplicial.*

Es preciso demostrar sólo la segunda parte de este teorema. Con este fin, a todo par de expresiones de tipo $(\lambda-\rho_i)^{m_i}$, $(\lambda+\rho_i)^{m_i}$, y en particular al par con $\rho_i=0$, ponemos en correspondencia el par de matrices

$$G_i=\begin{bmatrix} O & E_i \\ -E_i & O \end{bmatrix} \text{ y } A_i=\begin{bmatrix} B_i & O \\ O & -B_i \end{bmatrix},$$

donde E_i es la matriz unidad de orden m_i y B_i es la célula de

Jordan de orden m_i y de valor propio ρ_i. A las expresiones que quedan de tipo λ^{m_i} con m_i par ponemos en correspondencia el par de matrices

$$G_i = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & -1 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 1 & \dots & 0 & 0 \\ -1 & 0 & \dots & 0 & 0 \end{bmatrix} \quad \text{y} \quad A_i = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ & 0 & 1 & \dots & 0 & 0 \\ & & \dots & \dots & \dots & \dots \\ & & & & 0 & 1 \\ & & & & & 0 \end{bmatrix}.$$

Sea

$$G = G_1 \dotplus \dots \dotplus G_s \quad \text{y} \quad A = A_1 \dotplus \dots \dotplus A_s.$$

Consideremos un espacio bilineal métrico complejo $\mathfrak{L}$ de matriz de Gram igual a G. La matriz G es antisimétrica y regular y por esto el espacio $\mathfrak{L}$ es simplicial. La aplicación lineal $\mathcal{A}$ de matriz A tiene los divisores elementales dados. Al mismo tiempo, la aplicación $\mathcal{A}$ es antisimétrica, ya que los cálculos directos muestran que la matriz

$$AG = A_1G_1 \dotplus A_2G_2 \dotplus \dots \dotplus A_sG_s$$

es simétrica. Hemos demostrado el teorema.

27.3. Aplicaciones simpliciales. Las aplicaciones isométricas de un espacio simplicial $\mathfrak{L}$ suelen llamarse aplicaciones *simpliciales* de este espacio. Por consiguiente, si $\mathcal{U}$ es una aplicación simplicial del espacio $\mathfrak{L}$, se tiene $(x\mathcal{U}, y\mathcal{U}) = (x, y)$ para todos los x e y de $\mathfrak{L}$. En términos de matrices esta igualdad da

$$[x]\, UGU'\, [y]' = [x]\, G\, [y]',$$

de donde resulta

$$UGU' = G, \tag{7}$$

donde G es la matriz de Gram y U es la matriz de la aplicación $\mathcal{U}$. Tomando en $\mathfrak{L}$ un sistema simplicial de coordenadas, convertiremos la relación (7) en

$$USU' = S, \tag{8}$$

donde S es una matriz simplicial de tipo (1). Las matrices U que satisfacen la condición (8) serán llamadas matrices *simpliciales*. Por consiguiente, para que una aplicación lineal de un espacio simplicial sea simplicial es necesario y suficiente que su matriz sea simplicial en un sistema simplicial de coordenadas. De (8) también se desprende directamente que una suma directa de matrices simpliciales es una matriz simplicial.

Sea $\mathcal{U}$ una aplicación simplicial arbitraria de un espacio simplicial $\mathfrak{L}$. Representando $\mathfrak{L}$ en la forma de la suma directa de los subespacios radicales de la aplicación $\mathcal{U}$ y agrupando después los sumandos correspondientes a los valores propios recíprocamente

inversos de la aplicación $\mathcal{U}$, obtenemos la descomposición

$$\mathfrak{L} = \mathfrak{M}_{-1} + \mathfrak{M}_1 + \mathfrak{M}_2 + \ldots + \mathfrak{M}_s,$$

donde $\mathfrak{M}_{-1}$ y $\mathfrak{M}_1$ son los subespacios radicales correspondientes a los valores propios -1 y $+1$. Los subespacios $\mathfrak{M}_{-1}, \ldots, \mathfrak{M}_s$ son, en virtud del teorema 7 del p. 25.2, recíprocamente ortogonales.

El estudio del comportamiento de $\mathcal{U}$ sobre $\mathfrak{L}$ se reduce ahora al estudio del comportamiento de $\mathcal{U}$ sobre cada uno de los subespacios $\mathfrak{M}_i$ por separado. Repitiendo los razonamientos del p. 26.3 y empleando, en lugar del teorema sobre los divisores elementales de las aplicaciones antisimétricas de los espacios euclídeos complejos, el respectivo teorema para los espacios simpliciales, obtenemos la siguiente proposición:

TEOREMA 3. *Los divisores elementales de las aplicaciones simpliciales correspondientes a los valores propios diferentes de* ± 1 *aparecen en forma de pares* $(\lambda-\rho)^m$, $(\lambda-\rho^{-1})^m$; *los divisores elementales de tipo* $(\lambda \pm 1)^{2m+1}$ *también aparecen en forma de pares* $(\lambda+1)^{2m+1}$, $(\lambda+1)^{2m+1}$ *o* $(\lambda-1)^{2m+1}$, $(\lambda-1)^{2m+1}$ *y los divisores elementales de tipo* $(\lambda \pm 1)^{2m}$ *pueden aparecer un número arbitrario de veces. Recíprocamente, todo sistema de expresiones de tipo* $(\lambda-\rho_i)^{m_i}$, $\rho_i \neq 0$, *que posee estas propiedades es un sistema de divisores elementales de una aplicación simplicial.*

Las aplicaciones simpliciales que tienen divisores elementales iguales son, en virtud del teorema 1 del p. 25.1, isomorfas. Por esto el teorema 3 ofrece una clasificación completa, salvo un isomorfismo, de las aplicaciones simpliciales. Este teorema muestra, en particular, que si -1 es un valor propio de una aplicación simplicial, la multiplicidad de este valor propio es necesariamente par. Como todos los demás valores propios o bien son iguales a $+1$ o bien figuran en forma de pares recíprocamente inversos y de la misma multiplicidad, el producto de todos los valores propios de una aplicación simplicial es igual a $+1$ [1]. Este producto es igual al determinante de la matriz de la aplicación y llegamos así a la conclusión de que *los determinantes de las matrices de las aplicaciones simpliciales son iguales a la unidad.*

Ejemplos y problemas

1. Constrúyase una matriz simplicial de divisores elementales $\lambda+1$, $\lambda+1$ y $(\lambda-1)^2$.

2. Demuéstrese que las matrices de tipo $\begin{bmatrix} A & B \\ C & A' \end{bmatrix}$, donde B y C son matrices cuadradas antisimétricas de un mismo orden, contienen todo divisor elemental un número par de veces.

3. Enúnciese el teorema 1 como el teorema de pares de formas bilineales antisimétricas.

[1] Todo valor propio se cuenta tantas veces como indique su multiplicidad.

§ 28. Espacios seudounitarios

En el p. 24.2 hemos llamado espacios seudounitarios los espacios bilineales métricos hermitianos sobre el cuerpo de los números complejos provistos de una métrica no degenerada y simétrica: $(x, y) = \overline{(y, x)}$. También hemos demostrado en esa ocasión que todos estos espacios se determinan, salvo un isomorfismo, por su dimensión n y su signatura s y que, para un valor dado de n, s puede tomar los valores $n, n-2, \ldots, -n$. Al igual que antes, nos interesarán las propiedades de las aplicaciones simétricas, antisimétricas e isométricas. La clasificación de estas aplicaciones en el caso de espacios euclídeos complejos y de espacios simpliciales se basa en el teorema 1 del p. 25.1. En los espacios seudounitarios este teorema no tiene, en general, lugar. Consideremos, por ejemplo, un espacio bilineal métrico hermitiano de dos dimensiones $\mathfrak{L}$ de base e_1, y e_2 y de matriz de Gram $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.

Las aplicaciones lineales $\mathcal{U}_1$ y $\mathcal{U}_2$ definidas por las igualdades

$$e_1\mathcal{U}_1 = 2e_1, \quad e_1\mathcal{U}_2 = \frac{1}{2}e_1,$$

$$e_2\mathcal{U}_1 = \frac{1}{2}e_2, \quad e_2\mathcal{U}_2 = 2e_2$$

serán, obviamente, aplicaciones isométricas de este espacio. Los divisores elementales de las aplicaciones $\mathcal{U}_1$ y $\mathcal{U}_2$ son los mismos, a saber, $\lambda - 2$ y $\lambda - \frac{1}{2}$. Sin embargo, $\mathcal{U}_1$ y $\mathcal{U}_2$ no son isomorfas. En efecto, si las aplicaciones $\mathcal{U}_1$ y $\mathcal{U}_2$ fuesen isomorfas, los subespacios radicales correspondientes al valor propio 2 podrían ser transformados uno en el otro mediante un automorfismo del espacio $\mathfrak{L}$. El subespacio radical correspondiente al valor propio 2 de la aplicación $\mathcal{U}_1$ es la recta αe_1 y el subespacio análogo de la aplicación $\mathcal{U}_2$ es la recta αe_2. Al mismo tiempo, todos los vectores no nulos de la primera recta tienen un cuadrado positivo y todos los vectores no nulos de la segunda recta tienen un cuadrado negativo. Por consiguiente, ningún automorfismo del espacio $\mathfrak{L}$ puede transformar la primera recta en la segunda que es lo que se quería demostrar.

Este ejemplo muestra que la vía puramente geométrica para el estudio de las aplicaciones de los espacios seudounitarios resulta, en cierta medida, necesaria.

28.1. Aplicaciones simétricas. Sea $\mathfrak{L}$ un espacio seudounitario de dimensión n y de signatura s. Consideremos un subespacio lineal arbitrario $\mathfrak{N}$ de $\mathfrak{L}$. Convendremos en decir que un sistema de coordenadas de $\mathfrak{N}$ es *normal positivo* si su matriz de Gram es

de la forma

$$N = \begin{bmatrix} 0 & \dots & 0 & 1 \\ 0 & \dots & 1 & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 1 & \dots & 0 & 0 \end{bmatrix} \tag{1}$$

y que es *normal negativo* si su matriz de Gram es de la forma

$$N = \begin{bmatrix} 0 & \dots & 0 & -1 \\ 0 & \dots & -1 & 0 \\ \cdot & \cdot & \cdot & \cdot \\ -1 & \dots & 0 & 0 \end{bmatrix}. \tag{2}$$

Es obvio que no todo subespacio admite un sistema normal de coordenadas. Esto depende de la signatura del subespacio y de su dimensión. Los subespacios de dimensión par que admiten un sistema normal de coordenadas, positivo o negativo, son de signatura cero; los subespacios de dimensión impar con un sistema normal positivo de coordenadas son de signatura $+1$ y los que poseen un sistema normal negativo de coordenadas son de signatura -1. La demostración se deduce fácilmente de la definición de la signatura (p. 24.2).

Estudiemos los divisores elementales de una aplicación simétrica $\mathcal{A}$ de un espacio seudounitario $\mathfrak{L}$. Descompongamos $\mathfrak{L}$ en la suma directa de los subespacios radicales de la aplicación $\mathcal{A}$ y agrupemos los sumandos que corresponden a los valores propios conjugados. Así obtendremos la descomposición de $\mathfrak{L}$ en la suma directa de subespacios invariantes

$$\mathfrak{L} = \mathfrak{M}_1 \dotplus \mathfrak{M}_2 \dotplus \dots \dotplus \mathfrak{M}_t. \tag{3}$$

Estos subespacios son, según el p. 25.2, recíprocamente ortogonales y por ello el estudio de la aplicación $\mathcal{A}$ se reduce al estudio de la aplicación $\mathcal{A}$ en cada uno de los $\mathfrak{M}_i$. El espacio $\mathfrak{L}$ es no degenerado y por esto todos los $\mathfrak{M}_i$ son también no degenerados. Observemos ahora que los subespacios $\mathfrak{M}_i$ pueden ser de dos tipos: 1) $\mathfrak{M}_i$ es la suma directa de dos subespacios radicales correspondientes a valores propios conjugados no reales[1] y 2) $\mathfrak{M}_i$ es el subespacio radical correspondiente a un valor propio real. Consideremos el primer caso: sea

$$\mathfrak{M}_i = \mathfrak{M}_i' \dotplus \mathfrak{M}_i'',$$

[1] Si α es un valor propio y $\bar{\alpha}$ no es un valor propio, aceptaremos formalmente que $\mathfrak{L}_{\bar{\alpha}}$ es el subespacio nulo. De los razonamientos ulteriores se deduce, sin embargo, que este caso no puede darse.

donde $\mathfrak{M}_i'$ y $\mathfrak{M}_i''$ son los subespacios radicales correspondientes a los valores propios α y $\bar{\alpha}$, donde $\alpha \neq \bar{\alpha}$. Razonando igual que en el p. 25.1, obtenemos

$$(a_1', a_2') = 0 \quad \text{y} \quad (a_1'', a_2'') = 0, \tag{4}$$

donde a_1' y a_2' son vectores cualesquiera de $\mathfrak{M}_i'$ y a_1'' y a_2'' son vectores cualesquiera de $\mathfrak{M}_i''$. Si existe un vector a' de $\mathfrak{M}_i'$ ortogonal a todos los vectores de $\mathfrak{M}_i'$, el vector a' es isótropo en $\mathfrak{M}_i$. Como el subespacio $\mathfrak{M}_i$ es no degenerado, esto da $a' = o$. Luego, para todo vector no nulo del subespacio $\mathfrak{M}_i'$ existe un vector no ortogonal del subespacio $\mathfrak{M}_i''$.

Hemos supuesto que $\mathfrak{M}_i'$ es el subespacio radical de la aplicación $\mathcal{A}$ correspondiente a la raíz α. Por consiguiente, tenemos para un valor natural de m

$$\mathfrak{M}_i'(\mathcal{A} - \alpha\mathcal{E})^m = o \quad \text{y} \quad \mathfrak{M}_i'(\mathcal{A} - \alpha\mathcal{E})^{m-1} \neq o. \tag{5}$$

Tomemos en $\mathfrak{M}_i'$ un vector a tal que $a(\mathcal{A} - \alpha\mathcal{E})^{m-1} \neq o$. Puesto que $a(\mathcal{A} - \alpha\mathcal{E})^{m-1}$ es diferente de cero, en $\mathfrak{M}_i''$ existe un vector b no ortogonal a $a(\mathcal{A} - \alpha\mathcal{E})^{m-1}$. Normalicemos b de modo que

$$(a(\mathcal{A} - \alpha\mathcal{E})^{m-1}, b) = 1.$$

Pongamos ahora

$$a(\mathcal{A} - \alpha\mathcal{E})^j = a_{j+1} \quad \text{y} \quad b(\mathcal{A} - \bar{\alpha}\mathcal{E})^j = b_{j+1}$$
$$(j = 0, 1, \ldots, m-1).$$

Como la aplicación $\mathcal{A}$ es simétrica, resulta

$$(\mathcal{A} - \alpha\mathcal{E})^* = \mathcal{A} - \bar{\alpha}\mathcal{E} \quad \text{y} \quad (\mathcal{A} - \alpha\mathcal{E})^{*t} = (\mathcal{A} - \bar{\alpha}\mathcal{E})^t.$$

Por consiguiente, para $j + k = m + 1$ se tiene

$$(a_j, b_k) = (a(\mathcal{A} - \alpha\mathcal{E})^{j-1}, b(\mathcal{A} - \bar{\alpha}\mathcal{E})^{k-1}) =$$
$$= (a(\mathcal{A} - \alpha\mathcal{E})^{j+k-2}, b) = 1 \tag{6}$$

y para $j + k > m + 1$ se tiene

$$(a_j, b_k) = (a(\mathcal{A} - \alpha\mathcal{E})^{j+k-2}, b) = 0. \tag{7}$$

En particular, de (6) se deduce que $b(\mathcal{A} - \bar{\alpha}\mathcal{E})^{m-1} \neq o$. Además, si a' pertenece a $\mathfrak{M}_i'$, tenemos

$$(a', b(\mathcal{A} - \bar{\alpha}\mathcal{E})^m) = (a'(\mathcal{A} - \alpha\mathcal{E})^m, b) = 0.$$

En otras palabras el vector $b(\mathcal{A} - \bar{\alpha}\mathcal{E})^m$ del subespacio $\mathfrak{M}_i''$ es ortogonal a todos los vectores de $\mathfrak{M}_i'$. En virtud de la observación hecha, esto da $b(\mathcal{A} - \bar{\alpha}\mathcal{E})^m = o$. Indiquemos por $\mathfrak{N}_1'$ y $\mathfrak{N}_1''$ los subespacios tendidos sobre los vectores $a_1, a_2, \ldots, a_m$ y $b_1, b_2, \ldots, b_m$, respectivamente.

Sea

$$c_1 = a_1 + \alpha_2 a_2 + \ldots + \alpha_m a_m,$$
$$c_{j+1} = c_j(\mathcal{A} - \alpha\mathcal{E}) \quad (j = 1, \ldots, m-1).$$

De las fórmulas (6) y (7) se deduce que para cualesquiera valores de $\alpha_2, \ldots, \alpha_m$ se tiene $(c_1(\mathcal{A} - \alpha\mathcal{E})^{m-1}, b_1) = 1$, de donde obtenemos de nuevo

$$\begin{aligned} &(c_j, b_{m+1-j}) = 1 \quad \text{y} \quad (c_j, b_k) = 0 \\ &(j + k > m + 1; \ j, k = 1, \ldots, m). \end{aligned} \tag{8}$$

Es fácil ver que los coeficientes $\alpha_2, \ldots, \alpha_m$ se pueden escoger de manera que se cumplan las condiciones complementarias

$$(c_1, b_{m-1}) = (c_2, b_{m-2}) = \ldots = (c_{m-1}, b_1) = 0.$$

Entonces para $j + k < m + 1$ tendremos

$$(c_j, b_k) = (c_1, b_1(\mathcal{A} - \overline{\alpha}\mathcal{E})^{j+k-2}) = (c_1, b_{j+k-1}) = 0. \tag{9}$$

Sea $\mathfrak{N}_1 = \mathfrak{N}_1' + \mathfrak{N}_1''$. Las relaciones (4) (8) y (9) muestran que el sistema $c_1, \ldots, c_m, b_1, \ldots, b_m$ es un sistema *normal positivo* de coordenadas en $\mathfrak{N}_1$. Al mismo tiempo, la matriz de la aplicación $\mathcal{A}$ en este sistema de coordenadas se descompone en un par de células de Jordan de divisores elementales $(\lambda - \alpha)^m$ y $(\lambda - \overline{\alpha})^m$.

El subespacio $\mathfrak{N}_1$ es no degenerado; por consiguiente,

$$\mathfrak{M}_l = \mathfrak{N}_1 \dotplus \mathfrak{N}_1^{\perp},$$

donde $\mathfrak{N}_1^{\perp}$ es de nuevo un subespacio invariante respecto a $\mathcal{A}$. Si $\mathfrak{N}_1^{\perp} \neq o$, podemos, aplicando a $\mathfrak{N}_1^{\perp}$ el proceso expuesto, despejar en él un subespacio invariante $\mathfrak{N}_2$, etc. Así obtendremos para $\mathfrak{M}_l$ una representación de la forma

$$\mathfrak{M}_l = \mathfrak{N}_1 \dotplus \mathfrak{N}_2 \dotplus \ldots \dotplus \mathfrak{N}_{s_l},$$

donde $\mathfrak{N}_1, \ldots, \mathfrak{N}_{s_l}$ son subespacios invariantes recíprocamente ortogonales y en cada uno de ellos existe un sistema normal positivo de coordenadas en el que la matriz de la aplicación $\mathcal{A}$ se descompone en dos células de Jordan de divisores elementales $(\lambda - \alpha)^m$ y $(\lambda - \overline{\alpha})^m$.

Consideremos el segundo caso. Sea $\mathfrak{M}_l$ el subespacio radical correspondiente a un valor propio real α. Buscamos de nuevo un número natural m tal que

$$\mathfrak{M}_l(\mathcal{A} - \alpha\mathcal{E})^m = o \ \text{y} \ \mathfrak{M}_l(\mathcal{A} - \alpha\mathcal{E})^{m-1} \neq o.$$

Puesto que $(\mathcal{A} - \alpha\mathcal{E})^* = \mathcal{A} - \overline{\alpha}\mathcal{E} = \mathcal{A} - \alpha\mathcal{E}$, la aplicación $\mathcal{A} - \alpha\mathcal{E}$ y, con ella, también la aplicación $(\mathcal{A} - \alpha\mathcal{E})^{m-1}$ son simétricas. La correspondiente función bilineal hermitiana $(x(\mathcal{A} - \alpha\mathcal{E})^{m-1}, y)$ es también simétrica y no se anula idénticamente sobre $\mathfrak{M}_l$. Por

esto en $\mathfrak{M}_t$ existe un vector a tal que

$$(a\,(\mathcal{A}-\alpha\mathcal{E})^{m-1},\ a)=\beta\neq 0. \tag{10}$$

De la simetría de la función se deduce que β es real. Tomando $a_1=\sqrt{|\beta|^{-1}}\,a$, obtendremos en lugar de (10) la relación

$$(a_1\,(\mathcal{A}-\alpha\mathcal{E})^{m-1},\ a_1)=\varepsilon,\quad \varepsilon=\pm 1. \tag{11}$$

En esta relación el signo de la unidad depende de las propiedades de la misma aplicación $\mathcal{A}$ y, en todo caso, no lo podemos cambiar mediante la normalización. Pongamos

$$a_{j+1}=a_j(\mathcal{A}-\alpha\mathcal{E})\quad (j=1,\ \ldots,\ m-1).$$

De (11) resulta:

$$(a_j,\ a_{m+1-j})=(a_1\,(\mathcal{A}-\alpha\mathcal{E})^{m-1},\ a_1)=\varepsilon. \tag{12}$$

Análogamente, de la condición $\mathfrak{M}_t(\mathcal{A}-\alpha\mathcal{E})^m=o$ obtenemos

$$(a_j,\ a_k)=(a_1\,(\mathcal{A}-\alpha\mathcal{E})^{j+k-2},\ a_1)=0\quad (j+k>m+1). \tag{13}$$

Sea

$$b_1=a_1+\alpha_2a_2+\ldots+\alpha_ma_m,$$

donde $\alpha_2,\ \ldots,\ \alpha_m$ son, por ahora, unos números arbitrarios. Pongamos

$$b_{j+1}=b_j(\mathcal{A}-\alpha\mathcal{E})\quad (j=1,\ \ldots,\ m-1). \tag{14}$$

De las igualdades (12) y (13) resulta:

$$\begin{gathered}(b_j,\ b_{m+1-j})=\varepsilon\quad\text{y}\quad (b_j,\ b_k)=0\\ (j+k>m+1;\quad j,\ k=1,\ \ldots,\ m).\end{gathered} \tag{15}$$

Es fácil ver que los números $\alpha_2,\ \ldots,\ \alpha_m$ se pueden escoger de modo que se cumplan las relaciones

$$(b_1,\ b_{m-1})=(b_1,\ b_{m-2})=\ldots=(b_1,\ b_1)=0.$$

Entonces para $j+k<m+1$ tendremos

$$(b_j,\ b_k)=(b_1,\ b_1\,(\mathcal{A}-\alpha\mathcal{E})^{j+k-2})=(b_1,\ b_{j+k-1})=0. \tag{16}$$

Consideremos el subespacio $\mathfrak{N}_1$ tendido sobre $b_1,\ \ldots,\ b_m$. De (14) se deduce que $\mathfrak{N}_1$ es invariante respecto a $\mathcal{A}$ y que la matriz de la aplicación $\mathcal{A}$ en este sistema de coordenadas es igual a una célula de Jordan de divisor elemental $(\lambda-\alpha)^m$. Por otra parte, de (15) y (16) se ve que $b_1,\ \ldots,\ b_m$ es un sistema *normal* de coordenadas en $\mathfrak{N}_1$ y, además, *positivo*, si $\varepsilon=+1$, y *negativo*, si $\varepsilon=-1$. El espacio $\mathfrak{N}_1$ es no degenerado y, por ello, siendo $\mathfrak{N}_1\neq o$, se tiene

$$\mathfrak{M}_t=\mathfrak{N}_1\dotplus\mathfrak{N}_1^{\perp},$$

donde $\mathfrak{N}_1^{\perp}$ es un subespacio invariante de menor dimensión. Aplicando a $\mathfrak{N}_1^{\perp}$ el mismo proceso, despejaremos de $\mathfrak{N}_1^{\perp}$ un subespacio

invariante $\mathfrak{M}_2$, etc., hasta obtener la descomposición de $\mathfrak{M}_i$ en una suma directa de espacios recíprocamente ortogonales.

Aplicando ahora nuestros resultados a cada uno de los subespacios $\mathfrak{M}_i$, representaremos el espacio $\mathfrak{L}$ en la forma

$$\mathfrak{L} = \mathfrak{P}_1 \dot{+} \mathfrak{P}_2 \dot{+} \ldots \dot{+} \mathfrak{P}_l, \tag{17}$$

donde $\mathfrak{P}_1, \ldots, \mathfrak{P}_l$ son subespacios recíprocamente ortogonales, cada uno de los cuales corresponde o bien a un divisor elemental real $(\lambda-\alpha)^m$ de la aplicación $\mathcal{A}$ o bien es la suma directa de dos subespacios correspondientes a divisores elementales conjugados $(\lambda-\alpha)^m$ y $(\lambda-\overline{\alpha})^m$; además, en los subespacios del primer tipo existen sistemas normales, *positivos* o *negativos*, de coordenadas en los que la matriz de la aplicación $\mathcal{A}$ será una célula de Jordan, mientras que en los subespacios del segundo tipo existen sistemas normales *positivos* de coordenadas en los que la matriz de la aplicación $\mathcal{A}$ será la suma directa de dos células conjugadas de Jordan.

A todo subespacio $\mathfrak{P}_j$, correspondiente a una raíz real α, le corresponde un divisor elemental $(\lambda-\alpha)^m$ de la aplicación $\mathcal{A}$. Convengamos en poner en correspondencia al subespacio $\mathfrak{P}_j$ el divisor elemental $+(\lambda-\alpha)^m$, si en $\mathfrak{P}_j$ existe un sistema *normal positivo* de coordenadas, en el que la matriz de la aplicación $\mathcal{A}$ tiene la forma de una célula de Jordan, y el divisor elemental $-(\lambda-\alpha)^m$, si en $\mathfrak{P}_j$ existe un sistema *normal negativo* de coordenadas con la propiedad señalada. Mediante esta regla podemos, basándonos en la descomposición (17), obtener un sistema de divisores elementales de la aplicación $\mathcal{A}$ en el que todo divisor elemental real tendrá un signo determinado. Un sistema de divisores elementales de este tipo convendremos en denominarlo de *signo definido*. Repitiendo los razonamientos, realizados al final del p. 26.1, obtenemos que son isomorfas las aplicaciones simétricas del espacio $\mathfrak{L}$ que tienen unos sistemas de divisores elementales de signo definido iguales.

Supongamos ahora que tenemos dado un sistema arbitrario de expresiones de tipo $\pm(\lambda-\alpha_i)^{m_i}$, en el que todas las expresiones no reales aparecen en forma de pares conjugados y llevan el signo «más». ¿Existen un espacio seudounitario $\mathfrak{L}$ y una aplicación simétrica $\mathcal{A}$ del mismo tales que el sistema dado $\pm(\lambda-\alpha_i)^{m_i}$ sea un sistema de divisores elementales de signo definido de la aplicación $\mathcal{A}$? La respuesta es, obviamente, afirmativa. La construcción del espacio $\mathfrak{L}$ y de la aplicación $\mathcal{A}$ se puede realizar siguiendo el mismo método que ha sido empleado en el p. 26.1. Puesto que para esta construcción no se necesita nada nuevo, la dejamos a cargo del lector.

Queda por demostrar un resultado más sútil de que el sistema de divisores elementales de signo definido de una aplicación $\mathcal{A}$ no depende de cómo se escoja la descomposición (17) y se determina totalmente por la propia aplicación.

La descomposición (17) ha sido obtenida como resultado de la división de los subespacios $\mathfrak{M}_1, \ldots, \mathfrak{M}_t$ en subespacios cíclicos $\mathfrak{P}_i$. Puesto que los propios subespacios $\mathfrak{M}_j$ son o bien radicales o bien una suma de subespacios radicales, y, por consiguiente, se determinan unívocamente, la no unicidad puede surgir solamente al descomponer cada uno de los subespacios $\mathfrak{M}_j$. Los subespacios $\mathfrak{M}_j$ que son unas sumas de dos subespacios radicales no tienen interés para nosotros, ya que al descomponerlos se obtienen pares de subespacios cíclicos con pares de divisores elementales conjugados de tipo $(\lambda-\alpha)^m$ y $(\lambda-\overline{\alpha})^m$. Luego, resta demostrar solamente que como quiera que se descomponga un subespacio radical $\mathfrak{M}_j$ con raíz real α en una suma de subespacios no descomponibles y recíprocamente ortogonales, el sistema de divisores elementales de signo definido $\pm(\lambda-\alpha)^m$ será el mismo.

Sean

$$\mathfrak{M}_j = \mathfrak{N}_1 + \mathfrak{N}_2 + \ldots + \mathfrak{N}_t, \tag{18}$$

$$\mathfrak{M}_j = \mathfrak{P}_1 + \mathfrak{P}_2 + \ldots + \mathfrak{P}_t \tag{19}$$

dos descomposiciones de $\mathfrak{M}_j$ en sumas ortogonales de subespacios no descomponibles. Indiquemos por $\delta_i(\lambda-\alpha)^{m_i}$ y $\varepsilon_i(\lambda-\alpha)^{m_i}$ los divisores elementales de signo definido correspondientes a los subespacios $\mathfrak{N}_i$ y $\mathfrak{P}_i$, respectivamente ($i=1, \ldots, t$; $\delta_i, \varepsilon_i = \pm 1$). Coloquemos los sumandos de las sumas (18) y (19) de modo que $m_1 \geqslant m_2 \geqslant \ldots \geqslant m_t$. Supongamos que tenemos varios sumandos de dimensión mayor: $m_1 = \ldots = m_k > m_{k+1}$. Demostremos entonces que los sistemas de divisores elementales mayores $\varepsilon_1(\lambda-\alpha)^{m_1}, \ldots, \varepsilon_k(\lambda-\alpha)^{m_k}$ y $\delta_1(\lambda-\alpha)^{m_1}, \ldots, \delta_k(\lambda-\alpha)^{m_k}$ están compuestos de los mismos términos. Consideremos la función bilineal $f_1(x, y) = (x(\mathcal{A}-\alpha\mathcal{E})^{m_1-1}, y)$. La aplicación $\mathcal{A}-\alpha\mathcal{E}$ es simétrica y por esto la función $f_1(x, y)$ es también simétrica. Determinemos su signatura empleando la descomposición (18) y empleando la descomposición (19). Todos los subespacios de la descomposición (18) son ortogonales respecto a la función $f_1(x, y)$. Por ello la signatura de la función $f_1(x, y)$ sobre $\mathfrak{M}_j$ es igual a la suma de sus signaturas calculadas por separado sobre cada uno de los sumandos. Sin embargo, para $p > k$ tenemos

$$\mathfrak{N}_p(\mathcal{A}-\alpha\mathcal{E})^{m_1-1} = 0 .$$

y, por consiguiente, $f_1(x, y)$ se anula sobre $\mathfrak{N}_p$. Consideremos ahora los subespacios $\mathfrak{N}_1, \ldots, \mathfrak{N}_k$. Sea $a_1, \ldots, a_m$ una base de $\mathfrak{N}_1$ tal que

$$a_{j+1} = a_j(\mathcal{A}-\alpha\mathcal{E}) \qquad (j = 1, \ldots, m_1 - 1).$$

Puesto que $a_1(\mathcal{A}-\alpha\mathcal{E})^{m_1} = 0$, se tiene para $l+k > 2$

$$f_1(a_j, a_k) = (a_j(\mathcal{A}-\alpha\mathcal{E})^{m_1-1}, a_k) = 0. .$$

Por otra parte, según la definición del signo de un divisor elemental, tenemos

$$f_1(a_1, a_1) = (a_1(\mathcal{A}-\alpha\mathcal{E})^{m_1-1}, a_1) = \delta_1.$$

Por consiguiente, la matriz de la función $f_1(x, y)$ en la base $a_1, \ldots, a_m$ contendrá sólo un elemento no nulo que será igual a δ_1 y se hallará al principio de la diagonal principal. De aquí se ve que la signatura de la función $f_1(x, y)$ sobre $\mathfrak{N}_1$ también será igual a δ_1. Razonamientos análogos se pueden realizar también en el caso de los subespacios $\mathfrak{N}_2, \ldots, \mathfrak{N}_k$. Así obtenemos que la signatura de la función $f_1(x, y)$ sobre $\mathfrak{M}_j$ es igual a $\delta_1+\delta_2+\ldots+\delta_k$. Tomando en vez de la descomposición (18) la descomposición (19), llegaremos a la conclusión de que la signatura de la función $f_1(x, y)$ sobre $\mathfrak{M}_j$ es igual a $\varepsilon_1+\varepsilon_2+\ldots+\varepsilon_k$. Por consiguiente, se tiene

$$\delta_1+\delta_2+\ldots+\delta_k=\varepsilon_1+\varepsilon_2+\ldots+\varepsilon_k. \tag{20}$$

Puesto que $\delta_i, \varepsilon_i=\pm 1$, de (20) resulta que los sistemas de números $\delta_1, \ldots, \delta_k$ y $\varepsilon_1, \ldots, \varepsilon_k$ pueden diferir sólo en el orden de sucesión de los números y que, por consiguiente, ambos sistemas de divisores elementales mayores $\delta_1(\lambda-\alpha)^{m_1}, \ldots, \delta_k(\lambda-\alpha)^{m_k}$ y $\varepsilon_1(\lambda-\alpha)^{m_1}, \ldots, \varepsilon_k(\lambda-\alpha)^{m_k}$ coinciden.

Consideremos ahora la función bilineal

$$f_2(x, y)=(x(\mathcal{A}-\alpha\mathcal{E})^{m_{k+1}-1}, y).$$

Esta función se anula sobre todos los subespacios de dimensión menor que m_{k+1}. Por ello, para la signatura de $f_2(x, y)$ sobre $\mathfrak{M}_j$ tenemos las expresiones siguientes:

$$\left.\begin{aligned}\text{sign. }\mathfrak{M}_j&=\text{sign. }\mathfrak{N}_1+\ldots+\text{sign. }\mathfrak{N}_k+\text{sign. }\mathfrak{N}_{k+1}+\ldots+\text{sign. }\mathfrak{N}_p,\\ \text{sign. }\mathfrak{M}_j&=\text{sign. }\mathfrak{P}_1+\ldots+\text{sign. }\mathfrak{P}_k+\text{sign. }\mathfrak{P}_{k+1}+\ldots+\text{sign. }\mathfrak{P}_p,\end{aligned}\right\} \tag{21}$$

donde aceptamos que $m_{k+1}=\ldots=m_p>m_{p+1}$. Puesto que hemos demostrado ya que las signaturas de la función f_2 sobre los subespacios correspondientes de mayor dimensión coinciden, de (21) resulta que

$$\text{sign. }\mathfrak{N}_{k+1}+\ldots+\text{sign. }\mathfrak{N}_p=\text{sign. }\mathfrak{P}_{k+1}+\ldots+\text{sign. }\mathfrak{P}_p,$$

es decir,

$$\delta_{k+1}+\ldots+\delta_p=\varepsilon_{k+1}+\ldots+\varepsilon_p. \tag{22}$$

De (22) se ve que los divisores elementales de potencia m_{k+1} de la aplicación $\mathcal{A}$ también coinciden. Continuando este proceso, obtenemos que los sistemas de divisores elementales de signo definido de la aplicación $\mathcal{A}$, calculados mediante las descomposiciones (18) y (19), coinciden.

Veamos ahora qué puede decirse sobre la signatura del espacio principal $\mathfrak{L}$, si se conoce un sistema de divisores elementales de signo definido de una aplicación simétrica $\mathcal{A}$ de este espacio. La descomposición (17) muestra que la signatura del espacio $\mathfrak{L}$ es igual a la suma de las signaturas de los subespacios $\mathfrak{P}_1, \ldots, \mathfrak{P}_t$. En

aquellos subespacios $\mathfrak{P}_j$ que corresponden a pares conjugados de valores propios existe un sistema normal positivo de coordenadas. Sin embargo la dimensión de éstos es par y, por consiguiente, la signatura es igual a cero.

También es igual a cero la signatura de los subespacios que corresponden a divisores elementales reales de potencia par. Por ello quedan sólo los subespacios que corresponden a los divisores elementales reales de potencia impar. La signatura de éstos es igual a ± 1, según sea positivo o negativo el correspondiente divisor elemental. Por consiguiente,

$$s = s_1 - s_2, \tag{23}$$

donde s es la signatura del espacio $\mathfrak{L}$, s_1 es el número de divisores elementales positivos y s_2 el número de divisores elementales negativos reales de potencia impar de la aplicación $\mathcal{A}$.

De las fórmulas (23) se desprende, en particular, que *toda aplicación simétrica de un espacio seudounitario de signatura s tiene por lo menos* $|s|$ *divisores elementales reales de potencia impar.*

El estudio de las aplicaciones antisimétricas de los espacios seudounitarios se reduce directamente al estudio de las aplicaciones simétricas, ya que siendo $\mathcal{A}$ una aplicación simétrica, la aplicación $i\mathcal{A}$ será antisimétrica y viceversa.

28.2. Aplicaciones seudounitarias. Las aplicaciones isométricas de los espacios seudounitarios se llaman aplicaciones *seudounitarias*. Sea $\mathcal{U}$ una aplicación seudounitaria de un espacio $\mathfrak{L}$. Representando $\mathfrak{L}$ en la forma de la suma directa de los subespacios radicales de la aplicación $\mathcal{U}$ y agrupando los sumandos correspondientes a las raíces α y β, que cumplen la relación $\alpha\bar{\beta} = 1$, obtenemos una nueva descomposición de $\mathfrak{L}$:

$$\mathfrak{L} = \mathfrak{M}_{-1} + \mathfrak{M}_1 + \mathfrak{M}_2 + \ldots + \mathfrak{M}_s,$$

donde los subespacios $\mathfrak{M}_j$ son invariantes respecto a $\mathcal{U}$ y son, en virtud del teorema 3 del p. 25.1, recíprocamente ortogonales. A cada uno de estos subespacios se puede aplicar una de las fórmulas de Cayley (p. 25.2):

$$\mathcal{A} = i(\mathcal{E} - \mathcal{U})(\mathcal{E} + \mathcal{U})^{-1} \tag{24}$$

o

$$\mathcal{A} = i(\mathcal{E} + \mathcal{U})(\mathcal{E} - \mathcal{U})^{-1}. \tag{25}$$

Como resultado obtenemos unas aplicaciones simétricas de los subespacios $\mathfrak{M}_j$. De (24) y (25) tenemos

$$\mathcal{U} = (\mathcal{E} + i\mathcal{A})(\mathcal{E} - i\mathcal{A})^{-1}$$

y, respectivamente

$$\mathcal{U} = -(\mathcal{E} + i\mathcal{A})(\mathcal{E} - i\mathcal{A})^{-1}.$$

En el primer caso todo divisor elemental de la aplicación $\mathcal{A}$ de tipo $(\lambda-\alpha)^m$ se transforma en el divisor elemental $(\lambda-\gamma)^m$, donde $\gamma=(1+i\alpha)(1-i\alpha)^{-1}$, y en el segundo caso en el divisor elemental $(\lambda-\gamma)^m$, donde $\gamma=-(1+i\alpha)(1-i\alpha)^{-1}$. Puesto que $\mathcal{A}$ es una aplicación simétrica, sus divisores elementales no reales de tipo $(\lambda-\alpha)^m$ aparecen en forma de pares $(\lambda-\alpha)^m$, $(\lambda-\bar{\alpha})^m$. Sin embargo, $(1+i\bar{\alpha})(1-i\alpha)^{-1}=\overline{(1-i\alpha)(1+i\alpha)^{-1}}$ y, por ello, los correspondientes divisores elementales de la aplicación $\mathcal{U}$ también aparecerán en forma de pares, pero éstos serán ya de tipo $(\lambda-\gamma)^m$, $(\lambda-\bar{\gamma}^{-1})^m$. En cambio, si α es real, para $\gamma=(1+i\alpha)(1-i\alpha)^{-1}$ obtenemos la relación $\gamma\bar{\gamma}=1$. Recíprocamente, de $\gamma\bar{\gamma}=1$ se desprende que α es real. Por consiguiente, las fórmulas (24) y (25) establecen una correspondencia entre los divisores elementales de una aplicación seudounitaria $\mathcal{U}$ y los divisores elementales de una aplicación simétrica $\mathcal{A}$. Hemos visto que las aplicaciones simétricas se determinan, salvo un isomorfismo, por sus sistemas de divisores elementales de signo definido. Por esto las aplicaciones seudounitarias se determinan, salvo un isomorfismo, por sus sistemas de divisores elementales de signo definido. Pero en estos sistemas los signos «más» y «menos» deben ser colocados ahora ante los divisores elementales que corresponden a los valores propios de módulo igual a la unidad.

Ejemplos y problemas

1. Realícese la demostración completa de todas las afirmaciones de los p. p. 28.1 y 28.2.

2. Sea $\mathfrak{L}$ un espacio seudounitario de cuatro dimensiones de signatura 2 (espacio complejo de Lorentz). Tomemos como sistema principal de coordenadas de $\mathfrak{L}$ el sistema e_1, e_2, e_3 y e_4 en el que el cuadrado escalar de un vector $x=\xi_1e_1+\xi_2e_2+\xi_3e_3+\xi_4e_4$ es de la forma $(x, x)=\xi_1^2+\xi_2^2+\xi_3^2-\xi_4^2$. Demuéstrese que la matriz de toda aplicación simétrica del espacio $\mathfrak{L}$ puede ser reducida, en un adecuado sistema *principal* de coordenadas, a una de las formas siguientes:

$$\begin{bmatrix}\alpha & & & \\ & \beta & & \\ & & \gamma & \\ & & & \delta\end{bmatrix},\quad \begin{bmatrix}\alpha & & & \\ & \beta & & \\ & & \gamma & \delta \\ & & -\delta & \gamma\end{bmatrix}\quad y\quad \begin{bmatrix}\alpha & & 0 & 1 \\ & \beta & 1 & 0 \\ 0 & -1 & \beta & \\ -1 & 0 & & \beta\end{bmatrix},$$

donde α, β, γ y δ son números reales. Resuélvase el problema análogo para las matrices de las aplicaciones seudounitarias del espacio $\mathfrak{L}$.

3. Si n es la dimensión de un espacio y s es la signatura de una función bilineal hermitiana f, se llama *característica* de f la expresión $\chi=\frac{n-|s|}{2}$. Demuéstrese que la característica de un espacio seudounitario $\mathfrak{L}$, en el que actúa una aplicación simétrica $\mathcal{A}$, satisface la desigualdad

$$\chi\geqslant h+\sum\left[\frac{k}{2}\right],$$

donde h es igual a la mitad de la suma de los exponentes de los divisores elementales complejos de la aplicación $\mathcal{A}$ y k recorre todos los valores de los ex-

ponentes de sus divisores elementales reales. El símbolo $\left[\frac{k}{2}\right]$ representa el mayor número entero que no pasa de $\frac{k}{2}$.

4. Sea, en las condiciones del problema anterior, $f(x, y) = (x\mathcal{A}, y)$. Demuéstrese que la característica de la función f cumple la desigualdad

$$\chi \geqslant h + \sum \left[\frac{k'}{2}\right] + \sum \left[\frac{k''-1}{2}\right],$$

donde h es la mitad de la suma de los exponentes de los divisores elementales complejos de la aplicación $\mathcal{A}$, k' recorre todos los valores de los exponentes de los divisores elementales correspondientes a las raíces reales no nulas y k'' recorre todos los valores de los exponentes de los divisores elementales correspondientes al valor propio nulo.

Capítulo VIII

Espacios afines

El concepto del espacio vectorial real de tres dimensiones surge de un modo natural al describir las propiedades principales del espacio físico corriente, siempre que en este espacio se haya destacado un punto como el origen de coordenadas. Puesto que ningún punto del espacio físico tiene ventaja alguna ante los demás, conviene tener, además del concepto de espacio vectorial, un modelo matemático que corresponda al concepto del espacio en el que no se hayan fijado de antemano puntos algunos. Ejemplos de estos modelos son el espacio afín y el espacio euclídeo puntual. En este capítulo serán expuestos los elementos de la teoría de estos espacios.

§ 29. Espacios afines generales

29.1. Axiomática. Sea dado un espacio vectorial $\mathfrak{L}$ sobre un cuerpo conmutativo K. Se llama *espacio afín* $\mathfrak{A}$ *sobre el espacio vectorial* $\mathfrak{L}$ un conjunto arbitrario tal que a todo par de elementos X e Y del mismo le corresponde un vector de $\mathfrak{L}$, indicado en lo sucesivo por $\overrightarrow{XY}$, con la condición de que se cumplen las exigencias siguientes:

A_1: *para todo* $X \in \mathfrak{A}$ *y todo* $v \in \mathfrak{L}$ *existe un elemento* $Y \in \mathfrak{A}$, *y sólo uno, que satisface la relación* $\overrightarrow{XY} = v$;

A_2: *para cualesquiera* X, Y *y* Z *de* $\mathfrak{A}$ *es válida la igualdad* $\overrightarrow{XY} + \overrightarrow{YZ} = \overrightarrow{XZ}$.

Los elementos de un espacio afín se llaman *puntos* del mismo y los vectores de $\mathfrak{L}$ se llaman *vectores libres* del espacio $\mathfrak{A}$.

Para cualesquiera $X \in \mathfrak{A}$ y $v \in \mathfrak{L}$ indicaremos por $X \cdot v$ aquel punto $Y \in \mathfrak{A}$ para el cual $\overrightarrow{XY} = v$. Según el axioma A_1 un punto así de $\mathfrak{A}$ existe y es único. De aquí, en particular, se desprende que para cualesquiera X, $Y \in \mathfrak{A}$ y $u \in \mathfrak{L}$ se tiene

$$X \cdot \overrightarrow{XY} = Y \quad \text{y} \quad \overrightarrow{X \cdot (X \cdot u)} = u. \tag{1}$$

En virtud del axioma A_3 tenemos $\overline{XX}+\overline{XY}=\overline{XY}$ y, por ello,

$$\overline{XX}=o \quad \text{y} \quad X\cdot o=X \quad (X\in\mathfrak{A} \quad \text{y} \quad o\in\mathfrak{L}).$$

En vista del mismo axioma se tiene $\overline{XY}+\overline{YX}=\overline{XX}$ y, por ello,

$$\overline{XY}=-\overline{YX}. \tag{2}$$

Finalmente, para cualesquiera X, O, B, C, $D\in\mathfrak{A}$ se tiene

$$(X\cdot\overline{OB})\cdot\overline{CD}=X\cdot(\overline{OB}+\overline{CD}). \tag{3}$$

En efecto, sea $X\cdot\overline{OB}=Y$ e $Y\cdot\overline{CD}=Z$. Debido a A_3 y a (1), se tiene

$$\overline{OB}=\overline{XY}, \quad \overline{CD}=\overline{YZ} \quad \text{y} \quad \overline{OB}+\overline{CD}=\overline{XZ}$$

y por esto

$$(X\cdot\overline{OB})\cdot\overline{CD}=Z=X\cdot\overline{XZ}=X\cdot(\overline{OB}+\overline{CD}).$$

Pongamos en correspondencia a todo par (O, B) de puntos del espacio $\mathfrak{A}$ la aplicación de $\mathfrak{A}$ en $\mathfrak{A}$, indicada por $\overline{OB}$ y definida mediante la fórmula

$$\overrightarrow{OB}\colon\ X\longrightarrow X\cdot\overline{OB} \quad (X\in\mathfrak{A}), \tag{4}$$

es decir, tomemos $X\cdot\overrightarrow{OB}=X\cdot\overline{OB}$. Las aplicaciones $\overrightarrow{OB}$ $(O, B\in\mathfrak{A})$ se llaman *traslaciones* (o también *desplazamientos* o *traslados*) del espacio $\mathfrak{A}$. La superposición de unas traslaciones $\overrightarrow{OB}$ y $\overrightarrow{CD}$ suele llamarse suma de estas traslaciones, es decir, se toma por definición que

$$X\cdot(\overrightarrow{OB}+\overrightarrow{CD})=(X\cdot\overrightarrow{OB})\cdot\overrightarrow{CD}=X\cdot(\overline{OB}+\overline{CD}). \tag{5}$$

El producto de un elemento $\lambda\in K$ por una traslación $\overrightarrow{OB}$ se define mediante la fórmula

$$X\cdot(\lambda\overrightarrow{OB})=X\cdot(\lambda\overline{OB}). \tag{6}$$

El conjunto de todas las traslaciones del espacio $\mathfrak{A}$, provisto de las operaciones de adición y de multiplicación por todos los $\lambda\in K$, se indicará por $L(\mathfrak{A})$. Las igualdades (5) y (6) muestran que la aplicación $\overrightarrow{OX}\longrightarrow\overrightarrow{OX}$ es un isomorfismo de $\mathfrak{L}$ sobre $L(\mathfrak{A})$ y por esto $L(\mathfrak{A})$ es *un espacio vectorial sobre* K.

Fijemos ahora en $\mathfrak{A}$ un punto cualquiera O. Entonces la aplicación $X\longrightarrow\overrightarrow{OX}$ $(X\in\mathfrak{A})$ establece una correspondencia biyectiva entre los puntos de $\mathfrak{A}$ y los vectores de $\mathfrak{L}$. El vector $\overrightarrow{OX}$ suele llamarse *radio vector* del punto X respecto al punto inicial O. El radio vector del punto desplazado $X\cdot\overrightarrow{BC}$ se expresa en términos

del radio vector del punto dado inicialmente X mediante la fórmula

$$\overrightarrow{O(X \cdot \overrightarrow{BC})} = \overrightarrow{OX} + (\overrightarrow{OC} - \overrightarrow{OB}). \tag{7}$$

Efectivamente, tomando $X \cdot \overrightarrow{BC} = Y$, obtenemos $\overrightarrow{XY} = \overrightarrow{BC}$ y, por consiguiente,

$$\overrightarrow{O(X \cdot \overrightarrow{BC})} = \overrightarrow{OY} = \overrightarrow{OX} + \overrightarrow{XY} = \overrightarrow{OX} + \overrightarrow{BC}.$$

Supongamos ahora dado un espacio vectorial $\mathfrak{L}$ sobre un cuerpo conmutativo K. Llamemos puntos a los vectores de $\mathfrak{L}$ y pongamos en correspondencia a todo par (u, v) de ellos el vector $\overrightarrow{uv} = v - u$. Está claro que en este caso las exigencias A_1 y A_2 se cumplirán sin duda alguna y de esta forma el propio espacio vectorial $\mathfrak{L}$ se convertirá en un espacio afín sobre sí mismo. Indicaremos este espacio afín por $A(\mathfrak{L})$. Puesto que la igualdad $\overrightarrow{xy} = u$ equivale a la relación $y - x = u$, obtenemos para las traslaciones del espacio $A(\mathfrak{L})$ la fórmula

$$x \cdot \overrightarrow{bc} = x + (c - b). \tag{8}$$

Desde el punto de vista puramente algebraico, es conveniente a veces considerar, además del concepto de *un espacio afín sobre un espacio vectorial*, el concepto de un espacio afín *sobre un cuerpo conmutativo* que surge del modo siguiente.

Indiquemos por A el conjunto de todos los puntos de un espacio afín $\mathfrak{A}$ sobre un espacio vectorial $\mathfrak{L}$ sobre un cuerpo conmutativo principal K. Introducimos en A unas operaciones nuevas $S(X, Y, Z)$ y $P_\lambda(X, Y)$ tomando por definición

$$S(X, Y, Z) = Z \cdot \overrightarrow{XY} \text{ y } P_\lambda(X, Y) = X \cdot (\lambda \overrightarrow{XY}) \tag{9}$$

y convirtiendo con ello $\mathfrak{A}$ en un álgebra nueva

$$\mathfrak{A}^* = (A;\ S,\ \{P_\lambda\}) \quad (\lambda \in K).$$

Es fácil comprobar que en el álgebra $\mathfrak{A}^*$ las operaciones S y P_λ poseen las propiedades siguientes:

A_1^*: *Para cualesquiera* $O, B, C, X \in \mathfrak{A}^*$ se tiene

$$S(B, C, S(O, B, X)) = S(O, C, X); \tag{10}$$

A_2^*: *Para cualesquiera* $O, B, X \in \mathfrak{A}^*$ y $\lambda \in K$ se tiene

$$S(O, B, P_\lambda(O, X)) = P_\lambda(B, S(O, B, X)); \tag{11}$$

A_3^*: *Fijamos en* $\mathfrak{A}^*$ *un punto cualquiera* O *y mediante las operaciones* S *y* P_λ *introducimos para los elementos de* $\mathfrak{A}^*$ *unas operaciones nuevas*

$$X +_O Y = S(O, X, Y) \text{ y } \lambda \cdot_O X = P_\lambda(O, X), \tag{12}$$

llamadas adición local y multiplicación local de los elementos de $\mathfrak{A}^*$ *por los números* λ *(en el punto O). Respecto a estas operaciones locales* $+_O$ *y* $\lambda \cdot_O$ *los elementos de* $\mathfrak{A}^*$ *forman un espacio vectorial sobre K que se indica en lo sucesivo por* $\mathfrak{L}_O$.

En efecto, hemos visto que la aplicación $X \to \overline{OX}$ $(X \in A)$ es una correspondencia biyectiva entre los puntos de A y los vectores de $\mathfrak{L}$. Pasemos, de acuerdo con esta correspondencia, las operaciones vectoriales de $\mathfrak{L}$ a A, es decir, tomemos por definición

$$X + Y = Z \Leftrightarrow \overline{OX} + \overline{OY} = \overline{OZ} \text{ y } \lambda X = Z \Leftrightarrow \lambda \overline{OX} = \overline{OZ}. \quad (13)$$

Entonces A se convierte en un espacio lineal sobre K y el punto O será el elemento nulo de este espacio. Puesto que

$$X \cdot \lambda \overline{AB} = Y \Leftrightarrow \overline{XY} = \lambda \overline{AB} \Leftrightarrow \overline{OY} - \overline{OX} = \lambda (\overline{OB} - \overline{OA}),$$

y puesto que de (13) se deduce que

$$\overline{OY} - \overline{OX} = \lambda (\overline{OB} - \overline{OA}) \Leftrightarrow Y - X = \lambda (B - A),$$

tenemos

$$X \cdot \lambda \overline{AB} = Y \Leftrightarrow Y = X + \lambda (B - A),$$

de donde

$$X \cdot \lambda \overline{AB} = X + \lambda (B - A) \quad (14)$$

y, por consiguiente,

$$S(X, Y, Z) = Y + Z - X,$$
$$P_\lambda (X, Y) = \lambda (Y - X) + X,$$
$$X +_O Y = X + Y \quad \text{y} \quad \lambda \cdot_O X = \lambda X.$$

Después de esto las afirmaciones A_1^*, A_2^* y A_3^* se hacen evidentes.

Se llama espacio afín sobre un cuerpo conmutativo K *un conjunto de elementos* A *provisto de una operación ternaria* $S(X, Y, Z)$ *y una serie de operaciones binarias* $P_\lambda (X, Y)$ $(\lambda \in K)$ *que cumplen sobre* A *las exigencias* A_1^*, A_2^* *y* A_3^*.

Un espacio afín sobre un cuerpo conmutativo K, al igual que un espacio vectorial sobre un cuerpo conmutativo K, es un *álgebra* (cuya signatura depende de K y es, en el caso general, infinita). Por esto para los espacios afines sobre un cuerpo conmutativo resultan definidos automáticamente una serie de conceptos como isomorfismo, endomorfismo, congruencia, etc.

Hemos visto como todo espacio afín $\mathfrak{A}$ sobre un espacio vectorial $\mathfrak{L}$ sobre un cuerpo conmutativo K puede ser convertido en un espacio afín $\mathfrak{A}^*$ sobre el cuerpo conmutativo K. Queremos demostrar ahora que de esta forma se puede obtener cualquier espacio afín sobre K.

Sea, pues, $\mathfrak{B} = (A, S, P_\lambda)$ un espacio afín sobre un cuerpo conmutativo K. Representando la identidad (10) en la forma

$$C +_B (B +_O X) = C +_O X \quad (15)$$

y tomando $B = O$, obtenemos

$$C +_O (O + X) = C +_O X.$$

Esta es una ecuación en el espacio vectorial $\mathfrak{L}_\mathfrak{B}$ y, por ello,

$$O +_O X = X, \tag{16}$$

es decir, el punto O es el cero del espacio $\mathfrak{L}_O$.

Para cualesquiera puntos fijos O y B la aplicación

$$X \to S(O, B, X) \quad (X \in \mathfrak{B})$$

del espacio $\mathfrak{B}$ en sí mismo se llamará traslación de $\mathfrak{B}$ y se indicará por $\overrightarrow{OB}$. Por consiguiente, tomamos por definición

$$X \cdot \overrightarrow{OB} = S(O, B, X) = X +_O B. \tag{17}$$

Representando la relación (16) en la forma $X \cdot \overrightarrow{OO} = X$, vemos que la traslación $\overrightarrow{OO}$ es la aplicación idéntica de $\mathfrak{B}$ sobre sí mismo y que, en particular, $\overrightarrow{OO} = \overrightarrow{XX}$ para cualesquiera O, $X \in \mathfrak{B}$.

Al igual que antes, llamamos suma $\overrightarrow{OB} + \overrightarrow{CD}$ de traslaciones la superposición de las mismas, es decir, tomamos

$$X \cdot (\overrightarrow{OB} + \overrightarrow{CD}) = (X \cdot \overrightarrow{OB}) \cdot \overrightarrow{CD} = S(C, D, S(O, B, X)). \tag{18}$$

La identidad (10) significa que para las traslaciones es válida *la regla del triángulo*

$$\overrightarrow{OB} + \overrightarrow{BC} = \overrightarrow{OC}, \tag{19}$$

de la cual se desprende, en particular, que $\overrightarrow{OB} + \overrightarrow{BO} = \overrightarrow{OO}$, es decir, que las traslaciones $\overrightarrow{OB}$ y $\overrightarrow{BO}$ son aplicaciones *recíprocamente inversas*.

De (15) y (17) obtenemos

$$X \cdot \overrightarrow{CD} = X +_C D = X +_C (C +_B (D -_B C)) = X +_B (D -_B C),$$

es decir,

$$\overrightarrow{CD} = \overrightarrow{B\,(D -_B C)}. \tag{20}$$

Combinando las fórmulas (20) y (19), llegamos a la relación

$$\overrightarrow{OB} + \overrightarrow{CD} = \overrightarrow{OB} + \overrightarrow{B\,(D -_B C)} = \overrightarrow{O\,(D -_B C)}$$

que muestra que *la suma de traslaciones es una traslación*.

Definamos, finalmente, la operación de multiplicación de los números $\lambda \in K$ por una traslación $\overrightarrow{OB}$, poniendo

$$\lambda \cdot \overrightarrow{OB} = \overrightarrow{O(\lambda \cdot_O B)} = \overrightarrow{OP_\lambda(O, B)},$$

es decir, aceptando que

$$X \cdot (\lambda \overrightarrow{OB}) = S(O, P_\lambda(O, B), X) = P_\lambda(X, X \cdot \overrightarrow{OB}). \qquad (21)$$

De la fórmula (21) se ve que siendo $\overrightarrow{OB} = \overrightarrow{CD}$, se tiene $\lambda \cdot \overrightarrow{OB} = \lambda \cdot \overrightarrow{CD}$, es decir, la operación de multiplicación de un número por una traslación está definida unívocamente.

TEOREMA 1. *El conjunto $L(\mathfrak{A}^*)$ de todas las traslaciones de un espacio afín $\mathfrak{A}^*$ sobre un cuerpo conmutativo K es un espacio vectorial sobre K respecto a las operaciones de adición de traslaciones y de multiplicación de un número por una traslación. El espacio $L(\mathfrak{A}^*)$ es isomorfo a cualquier de los subespacios $\mathfrak{L}_O$. Poniendo en correspondencia a todo par de puntos X, Y de $\mathfrak{A}^*$ la traslación $\overrightarrow{XY} \in L(\mathfrak{A}^*)$ convertimos $\mathfrak{A}^*$ en un espacio afín $\mathfrak{A}$ sobre el espacio vectorial $L(\mathfrak{A}^*)$. Si en el espacio $\mathfrak{A}$ se introducen mediante las fórmulas (9) las operaciones S y P_λ, se obtiene de nuevo el espacio $\mathfrak{A}^*$.*

Las dos últimas afirmaciones son evidentes, ya que, por ejemplo, las fórmulas (9), aplicadas al espacio $\mathfrak{A}$, se convierten simplemente en las fórmulas (17) y (21). Demostremos las dos primeras afirmaciones.

Tomemos un punto cualquiera $O \in \mathfrak{A}^*$ y consideremos la aplicación

$$A: X \to \overrightarrow{OX} \quad (X \in \mathfrak{L}_O).$$

Puesto que cualquier traslación puede ser representada en la forma $\overrightarrow{OX}$, resulta que A es una aplicación de $\mathfrak{L}_O$ sobre $L(\mathfrak{A}^*)$. Veamos si esta aplicación A conserva las operaciones de adición y de multiplicación por números. Tenemos

$$Z \cdot \overrightarrow{O(X +_O Y)} = Z +_O (X +_O Y) = (Z +_O X) +_O Y = Z \cdot (\overrightarrow{OX} + \overrightarrow{OY}),$$

es decir, $\overrightarrow{O(X +_O Y)} = \overrightarrow{OX} + \overrightarrow{OY}$ y, por consiguiente,

$$(X +_O Y) A = XA + YA.$$

Análogamente, de (21) obtenemos $(\lambda \cdot_O X) A = \lambda \cdot XA$. Por consiguiente, la aplicación A es un isomorfismo de $\mathfrak{L}_O$ sobre $L(\mathfrak{A}^*)$ respecto a las operaciones de adición y de multiplicación por números y, por ello, $L(\mathfrak{A}^*)$, así como $\mathfrak{L}_O$, es un espacio vectorial sobre K.

Planteemos, finalmente, la pregunta: ¿bajo qué condiciones serán isomorfos dos espacios afines?

El concepto de isomorfismo suele introducirse para las álgebras arbitrarias con una misma signatura. Los espacios afines sobre un cuerpo conmutativo K han sido definidos como álgebras cuya signatura depende del cuerpo conmutativo K. Por ello para los espacios afines sobre *un mismo cuerpo conmutativo* está definido también el concepto de isomorfismo. En cuanto a los espacios afines sobre un espacio vectorial dado, éstos no han sido definidos como unas álgebras de signatura fija, sino como unos sistemas de una estructura más compleja para los cuales no hemos introducido por ahora el concepto de isomorfismo.

Introduzcamos la definición siguiente.

Se llama *homomorfismo* de un espacio afín $\mathfrak{A}$, definido sobre un espacio vectorial $\mathfrak{L}$, en un espacio afín $\mathfrak{B}$, definido sobre el mismo espacio vectorial $\mathfrak{L}$, toda aplicación puntual φ: $\mathfrak{A} \to \mathfrak{B}$ para la cual

$$\overline{XY} = \overline{X^{\varphi}Y^{\varphi}} \quad (X,\ Y \in \mathfrak{A}). \tag{22}$$

Está claro que dicha condición equivale a la siguiente:

$$(X \cdot v)^{\varphi} = X^{\varphi} \cdot v \quad (X \in \mathfrak{A} \quad \text{y} \quad v \in \mathfrak{L}).$$

Se ve de (22) que la aplicación φ transforma distintos puntos en distintos y que siendo φ un homomorfismo de $\mathfrak{A}$ sobre $\mathfrak{B}$, la aplicación inversa φ^{-1} es un homomorfismo de $\mathfrak{B}$ sobre $\mathfrak{A}$.

Un homomorfismo de $\mathfrak{A}$ sobre $\mathfrak{B}$ se llama isomorfismo si la aplicación inversa es un homomorfismo de $\mathfrak{B}$ sobre $\mathfrak{A}$. Luego, *para los espacios afines sobre un espacio vectorial fijo los conceptos de homomorfismo y de isomorfismo son equivalentes.*

Un isomorfismo de un espacio afín $\mathfrak{A}$, definido sobre un espacio vectorial $\mathfrak{L}$, sobre sí mismo se llama *automorfismo* de $\mathfrak{A}$ sobre $\mathfrak{L}$.

TEOREMA 2. *Todos los espacios afines sobre un mismo espacio vectorial son isomorfos. Los automorfismos de un espacio afín sobre un espacio vectorial son las traslaciones del mismo.*

Sean $\mathfrak{A}$ y $\mathfrak{B}$ unos espacios afines sobre un espacio vectorial $\mathfrak{L}$. Tomemos unos puntos arbitrarios $A \in \mathfrak{A}$ y $B \in \mathfrak{B}$ y consideremos la aplicación φ: $Av \to Bv$ $(v \in \mathfrak{L})$. De los axiomas A_1 y A_2 deducimos directamente que φ es un isomorfismo de $\mathfrak{A}$ sobre $\mathfrak{B}$.

Además, si φ es un automorfismo del espacio afín $\mathfrak{A}$ sobre el espacio vectorial $\mathfrak{L}$, obtenemos de la fórmula (22) que $\overline{AX} = \overline{A^{\varphi}X^{\varphi}}$, de donde $X^{\varphi} = A^{\varphi} \cdot \overline{AX}$ y por ello

$$X^{\varphi} = X \cdot \overline{XA^{\varphi}} \cdot \overline{AX} = X \cdot (\overline{AX} + \overline{XA^{\varphi}}) = X \cdot \overline{AA^{\varphi}},$$

es decir, φ es la traslación de $\mathfrak{A}$ determinada por el vector $\overline{AA^{\varphi}}$. Recíprocamente, si φ es una traslación, se tiene que φ es de la forma $X^{\varphi} = X \cdot v$, donde v es un vector arbitrario fijo de $\mathfrak{L}$, y por lo tanto

$$\overline{X^{\varphi}Y^{\varphi}} = \overline{(X \cdot v)(Y \cdot v)} = \overline{XY}.$$

Un espacio afín sobre un cuerpo conmutativo K es un álgebra de signatura S, P_{λ} $(\lambda \in K)$.

Un isomorfismo φ de un espacio afín $\mathfrak{A}$, definido sobre un cuerpo conmutativo K, sobre un espacio afín $\mathfrak{B}$, definido sobre el mismo cuerpo K, es una aplicación biyectiva de $\mathfrak{A}$ sobre $\mathfrak{B}$ que satisface las condiciones

$$S(X, Y, Z)^{\varphi} = S(X^{\varphi}, Y^{\varphi}, Z^{\varphi}),$$
$$P_{\lambda}(X, Y)^{\varphi} = P_{\lambda}(X^{\varphi}, Y^{\varphi})$$

o, que es equivalente, las condiciones

$$(Z \cdot \overline{XY})^{\varphi} = Z^{\varphi} \cdot \overline{X^{\varphi}Y^{\varphi}}, \tag{23}$$
$$(X \cdot \lambda\overline{XY})^{\varphi} = X^{\varphi} \cdot \lambda\overline{X^{\varphi}Y^{\varphi}}. \tag{24}$$

TEOREMA 3. *Para que una aplicación* φ: $\mathfrak{A} \to \mathfrak{B}$ *de un espacio afín* $\mathfrak{A}$, *definido sobre un cuerpo conmutativo* K, *sobre un espacio afín* $\mathfrak{B}$, *definido sobre el mismo cuerpo, sea un isomorfismo, es necesario y suficiente que la aplicación* ψ: $\overline{XY} \to \overline{X^{\varphi}Y^{\varphi}}$ $(X, Y \in \mathfrak{A})$ *sea un isomorfismo del espacio vectorial* $L(\mathfrak{A})$ *sobre el espacio vectorial* $L(\mathfrak{B})$.

NECESIDAD. Probemos, ante todo, que la aplicación ψ es unívoca. Sea $\overline{XY} = \overline{UV}$ $(X, Y, U, V \in \mathfrak{A})$ y, por consiguiente, $Z \cdot \overline{XY} = Z \cdot \overline{UV}$. De (23) obtenemos $Z^{\varphi} \cdot \overline{X^{\varphi}Y^{\varphi}} = Z^{\varphi} \cdot \overline{U^{\varphi}V^{\varphi}}$. Puesto que Z^{φ} recorre todo el espacio $\mathfrak{B}$, de la última igualdad tenemos $\overline{X^{\varphi}Y^{\varphi}} = \overline{U^{\varphi}V^{\varphi}}$ y por esto la aplicación ψ es unívoca. Representando ahora la relación (23) en la forma $(Zu)^{\varphi} = Z^{\varphi}u^{\psi}$, obtenemos para cualesquiera v, $w \in L(\mathfrak{A})$

$$Z^{\varphi}(v+w)^{\psi} = (Z(v+w))^{\varphi} = ((Zv)w)^{\varphi} = (Z^{\varphi}v^{\psi})w^{\psi} = Z^{\varphi}(v^{\psi}+w^{\psi}),$$

de donde

$$(v+w)^{\psi} = v^{\psi} + w^{\psi}. \tag{25}$$

Análogamente, representando (24) en la forma $(X \cdot \lambda u)^{\varphi} = X^{\varphi} \cdot \lambda u^{\psi}$, obtenemos

$$X^{\varphi}(\lambda u)^{\psi} = (X \cdot (\lambda u))^{\varphi} = X^{\varphi} \cdot \lambda u^{\psi},$$

de donde

$$(\lambda u)^{\psi} = \lambda u^{\psi}. \tag{26}$$

Las relaciones (25) y (26) significan que ψ es un isomorfismo de $L(\mathfrak{A})$ sobre $L(\mathfrak{B})$.

SUFICIENCIA. Supongamos que dada la aplicación φ: $\mathfrak{A} \to \mathfrak{B}$ del conjunto $\mathfrak{A}$ sobre el conjunto $\mathfrak{B}$, la aplicación ψ: $\overline{XY} \to \overline{X^{\varphi}Y^{\varphi}}$ es un isomorfismo de $L(\mathfrak{A})$ sobre $L(\mathfrak{B})$. Puesto que en un isomorfismo de espacios vectoriales el vector nulo corresponde al vector nulo, de $X^{\varphi} = Y^{\varphi}$ resulta $\overline{X^{\varphi}Y^{\varphi}} = o$ y por esto $\overline{XY} = o$ y $X = Y$, es decir, la aplicación φ es biyectiva. Para demostrar ahora que para φ es válida la relación (24), tomemos en $\mathfrak{A}$ un punto U tal que $\overline{XY} = \overline{ZU}$.

Entonces se tiene

$$(Z\cdot\overrightarrow{XY})^{\varphi}=U^{\varphi}=Z^{\varphi}\cdot\overrightarrow{Z^{\varphi}U^{\varphi}}=Z^{\varphi}\cdot\overrightarrow{X^{\varphi}Y^{\varphi}} \tag{27}$$

y, en particular,

$$(Zu)^{\varphi}=Z^{\varphi}u^{\psi}\quad(u\in L(\mathfrak{A})).$$

Ahora resulta de (26) que

$$(X\cdot\lambda\overrightarrow{XY})^{\varphi}=X^{\varphi}\cdot(\lambda\overrightarrow{XY})^{\psi}=X^{\varphi}\lambda\overrightarrow{XY}^{\psi}=X^{\varphi}\lambda\overrightarrow{X^{\varphi}Y^{\varphi}}$$

que es lo que se quería demostrar.

COROLARIO 1. *Los espacios afines sobre un cuerpo conmutativo dado son isomorfos cuando, y sólo cuando, sus espacios vectoriales de traslaciones son de una misma dimensión.*

Efectivamente, según el teorema 3, un isomorfismo entre espacios afines sobre un cuerpo conmutativo equivale a un isomorfismo de los espacios vectoriales de traslaciones y un isomorfismo de espacios vectoriales equivale (véase el p. 4.3) a la coincidencia de sus dimensiones.

COROLARIO 2. *Los automorfismos de un espacio afín $\mathfrak{A}$ sobre un cuerpo conmutativo que dejan en su sitio un punto $O\in\mathfrak{A}$ son aplicaciones de tipo* φ: $X\longrightarrow O\cdot\overrightarrow{OX}^{\psi}$, *donde* ψ *es un automorfismo fijo cualquiera del espacio vectorial* $L(\mathfrak{A})$.

En efecto, siendo φ un automorfismo del espacio $\mathfrak{A}$ que deja en su sitio el punto O, se tiene

$$X^{\varphi}=O\cdot\overrightarrow{OX}^{\varphi}=O\cdot\overrightarrow{O^{\varphi}X^{\varphi}}=O\cdot\overrightarrow{OX}^{\psi}.$$

Recíprocamente, si ψ es un automorfismo de $L(\mathfrak{A})$, se tiene

$$\overrightarrow{X^{\varphi}Y^{\varphi}}=\overrightarrow{(O\cdot\overrightarrow{OX}^{\psi})\,(O\cdot\overrightarrow{OY}^{\psi})}=\overrightarrow{OY}^{\psi}-\overrightarrow{OX}^{\psi}=(\overrightarrow{OY}-\overrightarrow{OX})^{\psi}=\overrightarrow{XY}^{\psi}$$

y por esto φ es, según el teorema 3, un automorfismo del espacio $\mathfrak{A}$.

TEOREMA 4. *Siendo O un punto arbitrario de un espacio afín $\mathfrak{A}$ sobre un cuerpo conmutativo K, todo automorfismo φ del espacio $\mathfrak{A}$ sobre K puede ser representado en la forma*

$$\varphi=\varphi_O\varphi_L, \tag{28}$$

donde φ_L es una traslación adecuada de $\mathfrak{A}$ y φ_O es un automorfismo de $\mathfrak{A}$ que deja en su sitio el punto O. Para todo automorfismo φ la descomposición de tipo (28) *es unívoca.*

Sea φ un automorfismo del espacio $\mathfrak{A}$. Entonces la aplicación $\varphi_O=\varphi\cdot\overrightarrow{O^{\varphi}O}$ es un automorfismo de $\mathfrak{A}$ y de las relaciones

$$O\cdot\varphi_O=O^{\varphi}\cdot\overrightarrow{O^{\varphi}O}=O$$

se ve que φ_O deja en su sitio el punto O. Poniendo $\varphi_L=\overrightarrow{OO^{\varphi}}$, obte-

nemos la igualdad (28). Sea ahora

$$\varphi = \varphi_O \varphi_L = \varphi_O^{(1)} \varphi_L^{(1)},$$

donde $\varphi_L^{(1)}$ es una traslación y $\varphi_O^{(1)}$ un automorfismo de $\mathfrak{A}$ que deja en su sitio a O. Tenemos entonces

$$O \cdot \varphi_L^{(1)} \varphi_L^{-1} = O \cdot \varphi_O^{(1)} \varphi_O^{-1} = O.$$

Puesto que la traslación $\varphi_L^{(1)} \varphi_L^{-1}$ deja en su sitio el punto O, se tiene $\varphi_L^{(1)} \varphi_L^{-1} = \overrightarrow{OO}$, de donde $\varphi_L^{(1)} = \varphi_L$ y $\varphi_O^{(1)} = \varphi_O$.

29.2. Variedades lineales. Sea $\mathfrak{A}$ un espacio afín sobre un cuerpo conmutativo K. El espacio vectorial $L(\mathfrak{A})$ formado por las traslaciones de $\mathfrak{A}$ lo indicaremos por $\mathfrak{L}$. A todo par de puntos $X, Y \in \mathfrak{A}$ corresponderá entonces un vector $\overrightarrow{XY} = \overrightarrow{XY}$ de $\mathfrak{L}$ y esta correspondencia satisfará los axiomas A_1 y A_2 del p. 29.1. Se llama *dimensión* del espacio afín $\mathfrak{A}$ la dimensión del correspondiente espacio vectorial $\mathfrak{L}$. En particular, un espacio afín de dimensión 0 está compuesto por un punto solo.

Unos puntos $X_0, X_1, \ldots, X_m$ del espacio $\mathfrak{A}$ se llaman *linealmente independientes*, si son linealmente independientes los vectores $\overrightarrow{X_0X_1}, \overrightarrow{X_0X_2}, \ldots, \overrightarrow{X_0X_m}$. Se dice que un punto $X \in \mathfrak{A}$ *depende linealmente* de la sucesión de puntos $X_0, X_1, \ldots, X_m$, si el vector $\overrightarrow{X_0X}$ puede ser expresado linealmente en términos de los vectores $\overrightarrow{X_0X_1}, \ldots, \overrightarrow{X_0X_m}$.

En estas definiciones el punto X_0 desempeña un papel especial. Pero de hecho, la dependencia lineal no está ligada al orden de los puntos en la sucesión $X_0, X_1, \ldots, X_m$. Sea O un punto arbitrario del espacio $\mathfrak{A}$. Entonces se tiene

$$\overrightarrow{X_0X_i} = \overrightarrow{OX_i} - \overrightarrow{OX_0} \quad (i = 0, 1, \ldots, m)$$

y el hecho de que el sistema de puntos $X_0, \ldots, X_m$ sea linealmente independiente equivale a que para cualesquiera $\lambda_1, \ldots, \lambda_m \in K$ de la relación

$$\lambda_1 (\overrightarrow{OX_1} - \overrightarrow{OX_0}) + \ldots + \lambda_m (\overrightarrow{OX_m} - \overrightarrow{OX_0}) = o$$

se desprenda $\lambda_1 = \ldots = \lambda_m = 0$. Tomando $\lambda_0 = -\lambda_1 - \ldots - \lambda_m$, vemos que *los puntos $X_0, X_1, \ldots, X_m$ son linealmente independientes cuando, y sólo cuando, para cualesquiera $\lambda_0, \lambda_1, \ldots, \lambda_m \in K$ de las relaciones*

$$\lambda_0 + \lambda_1 + \ldots + \lambda_m = 0,$$
$$\lambda_0 \overrightarrow{OX_0} + \lambda_1 \overrightarrow{OX_1} + \ldots + \lambda_m \overrightarrow{OX_m} = o$$

se deduce que $\lambda_0 = \lambda_1 = \ldots = \lambda_m = 0$.

En esta afirmación no destaca ya ninguno de los puntos $X_0, X_1, \ldots, X_m$ y por esto si la sucesión de puntos $X_0, X_1, \ldots, X_m$ es linealmente dependiente, también será linealmente dependiente la sucesión $X_{i_0}, X_{i_1}, \ldots, X_{i_m}$ cualquiera que sea la permutación $(i_0, i_1, \ldots, i_m)$ de los números $0, 1, \ldots, m$. Esto también se refiere al concepto de dependencia lineal de un punto X respecto a una sucesión $X_0, X_1, \ldots, X_m$.

En las definiciones dadas se supone que la sucesión $X_0, X_1, \ldots, X_m$ tiene no menos de dos términos ($m \geqslant 1$). Por definición, se acepta que *un sistema compuesto de un punto es linealmente dependiente y que el hecho de que un punto X dependa linealmente del sistema X_0 significa que $X = X_0$.*

De los teoremas del p. 4.2 sobre la dependencia lineal de los vectores obtenemos directamente que

a) *una sucesión de puntos $X_0, X_1, \ldots, X_m$ ($m \geqslant 1$) es linealmente dependiente cuando, y sólo cuando, al menos uno de sus términos depende linealmente de los demás;*

b) *si un punto X depende linealmente de los puntos $X_0, \ldots, X_m$ y todo punto X_i depende linealmente de los puntos $Y_0, \ldots, Y_k$, el punto X depende linealmente de los puntos $Y_0, \ldots, Y_k$;*

c) *si un punto X depende linealmente de los puntos $X_1, \ldots, X_m$, el punto X depende linealmente también de los puntos $X_0, X_1, \ldots, X_m$, donde X_0 es un punto cualquiera del espacio.*

De la propiedad c) se deduce que si un sistema de puntos $X_0, X_1, \ldots, X_m$ es linealmente independiente, cualquier subsistema del mismo que contiene más de un punto es también linealmente independiente. Esto permite introducir la definición siguiente: un conjunto arbitrario (finito o infinito) de puntos $\mathfrak{M}$ de un espacio afín $\mathfrak{A}$ se llama *linealmente independiente*, si todo su subconjunto finito que contiene más de un punto es linealmente independiente.

Análogamente se dice que un punto $X \in \mathfrak{A}$ *depende linealmente* de un conjunto de puntos $\mathfrak{M}$, si X depende linealmente de algún subconjunto finito de puntos de $\mathfrak{M}$.

Se llama *adherencia lineal* de un conjunto cualquiera no vacío $\mathfrak{M}$ de puntos de un espacio afín $\mathfrak{A}$ el conjunto formado por todos los puntos $X \in \mathfrak{A}$ que dependen linealmente del conjunto $\mathfrak{M}$. Un conjunto $\mathfrak{M}$ se llama *linealmente cerrado* si $\mathfrak{M}$ coincide con su adherencia lineal.

Los conjuntos linealmente cerrados de puntos de un espacio afín se llaman *planos* o *variedades lineales* del mismo. En otras palabras, un conjunto $\mathfrak{M}$ de puntos de un espacio afín $\mathfrak{A}$ se llama plano de $\mathfrak{A}$, si cualesquiera que sean $X_1, \ldots, X_m \in \mathfrak{M}$ este conjunto contiene también cualquier punto del espacio $\mathfrak{A}$ que dependa linealmente de los puntos $X_1, \ldots, X_m$.

De aquí se deduce directamente que todo punto y el propio espacio $\mathfrak{A}$ son indudablemente planos.

Dos planos $\mathfrak{M}$ y $\mathfrak{N}$ se llaman *incidentes* si uno de ellos está contenido en el otro. Si $\mathfrak{M} \subseteq \mathfrak{N}$, se dice también que $\mathfrak{M}$ *se encuentra sobre* $\mathfrak{N}$ (o que $\mathfrak{N}$ *pasa por* $\mathfrak{M}$).

Entre los planos de un espacio afín $\mathfrak{A}$ y los subespacios lineales del espacio vectorial $L(\mathfrak{A})$ existe una relación muy simple que también se emplea a veces para definir el concepto de plano.

TEOREMA 1. *Si para un conjunto* $\mathfrak{M}$ *de puntos de un espacio afín* $\mathfrak{A}$ *y un punto* $P \in \mathfrak{M}$ *el conjunto* $L(\mathfrak{M})$ *de los vectores de tipo* $\overrightarrow{PX}$ $(X \in \mathfrak{M})$ *es un subespacio lineal del espacio vectorial* $L(\mathfrak{A})$, *el conjunto* $\mathfrak{M}$ *es un plano de* $\mathfrak{A}$. *Para cualquier plano no vacío* $\mathfrak{M}$ *de un espacio afín* $\mathfrak{A}$ *y para cualquier punto* $P \in \mathfrak{M}$ *el conjunto* $L(\mathfrak{M})$ *de todos los vectores de tipo* $\overrightarrow{PX}$ $(X \in \mathfrak{M})$ *es un subespacio lineal de* $L(\mathfrak{A})$ *que no depende de la selección del punto* P *en* $\mathfrak{M}$.

Efectivamente, sea $P \in \mathfrak{M}$ y sea $L(\mathfrak{M})$ un subespacio de $L(\mathfrak{A})$. Si $X_1, \ldots, X_m \in \mathfrak{M}$ y un punto $X \in \mathfrak{A}$ depende linealmente de $X_1, \ldots, X_m$, se tiene entonces para unos valores convenientes $\lambda_1, \ldots, \lambda_m \in K$

$$\overrightarrow{PX} = \lambda_1 \overrightarrow{PX_1} + \ldots + \lambda_m \overrightarrow{PX_m} \in L(\mathfrak{M}),$$

es decir, para un punto $Y \in \mathfrak{M}$ convenientemente escogido se tiene $\overrightarrow{PX} = \overrightarrow{PY}$, de donde $X = Y \in \mathfrak{M}$.

Recíprocamente, sea $\mathfrak{M}$ un plano de $\mathfrak{A}$ y sean $P, X, Y \in \mathfrak{M}$ y $\lambda, \mu \in K$. Poniendo

$$Q = P \cdot (\lambda \overrightarrow{PX} + \mu \overrightarrow{PY}),$$

vemos que

$$\overrightarrow{PQ} = \lambda \overrightarrow{PX} + \mu \overrightarrow{PY},$$

es decir, el punto Q depende linealmente de P, X e Y y, por consiguiente, $Q \in \mathfrak{M}$ y $\overrightarrow{PQ} \in L(\mathfrak{M})$; es decir, $L(\mathfrak{M})$ es un subespacio lineal de $L(\mathfrak{A})$. Para cualesquiera $P, P_1, X \in \mathfrak{M}$ tenemos $\overrightarrow{P_1X} = \overrightarrow{PX} - \overrightarrow{PP_1}$ y, por consiguiente, el conjunto de los vectores de tipo $\overrightarrow{P_1X}$ $(X \in \mathfrak{M})$ coincide con el conjunto de los vectores de tipo $\overrightarrow{PX}$ $(X \in \mathfrak{M})$.

Señalemos dos importantes corolarios que se desprenden del teorema 1.

COROLARIO 1. *Sea* O *un punto fijo de un espacio afín* $\mathfrak{A}$, *sea* $\mathfrak{M}$ *un plano que pasa por* O *y sea* P *un punto arbitrario de* $\mathfrak{A}$. *Entonces el conjunto* $\mathfrak{N} = \mathfrak{M} \cdot \overrightarrow{OP}$ *formado por todos los puntos desplazados del plano* $\mathfrak{M}$ *es un plano de* $\mathfrak{A}$ *que pasa por el punto* P. *Recíprocamente, si* $\mathfrak{N}$ *es un plano que pasa por* P, *el conjunto* $\mathfrak{M} = \mathfrak{N} \cdot \overrightarrow{OP}$ *es un plano que pasa por* O.

En efecto, de $O \in \mathfrak{M}$ se deduce que $P = O \cdot \overrightarrow{OP} \in \mathfrak{N}$ y para todo punto $X \in \mathfrak{N}$ existe un punto $Y \in \mathfrak{M}$ tal que $X = Y \cdot \overrightarrow{OP}$. De aquí se tiene $\overrightarrow{YX} = \overrightarrow{OP}$ y $\overrightarrow{PX} = \overrightarrow{OY}$ y por lo tanto $L(\mathfrak{M}) = L(\mathfrak{M} \cdot \overrightarrow{OP})$.

COROLARIO 2. *Para todo punto fijo O de un espacio afín $\mathfrak{A}$ la aplicación $X \longrightarrow \overrightarrow{OX}$ $(X \in \mathfrak{A})$ determina una correspondencia biyectiva entre los subespacios lineales del espacio vectorial $L(\mathfrak{A})$ y los planos del espacio $\mathfrak{A}$ que pasan por O.*

La demostración es evidente.

Hemos señalado en el p. 29.1 que si llamamos punto a todo vector v de un espacio vectorial $\mathfrak{L}$ y si tomamos $\overrightarrow{vw} = w - v$, se obtiene el espacio afín $A(\mathfrak{L})$ sobre $\mathfrak{L}$. ¿Qué conjuntos de vectores de $\mathfrak{L}$ serán los planos de $A(\mathfrak{L})$?

Tomando por el punto fijo O de $A(\mathfrak{L})$ el vector nulo, vemos que la aplicación $X \longrightarrow \overrightarrow{OX}$, de la cual se trata en el corolario 2, se convierte en la aplicación idéntica $w \longrightarrow w$. Luego, son planos del espacio $A(\mathfrak{L})$ que pasan por O todos los subespacios lineales del espacio $\mathfrak{L}$, y sólo ellos. Los planos que pasan por un punto arbitrario $p \in \mathfrak{L}$ son, según el corolario 1, los conjuntos de vectores de tipo $p + \mathfrak{M}$, donde $\mathfrak{M}$ es un subespacio lineal del espacio $\mathfrak{L}$.

Se llama *base* de un plano arbitrario $\mathfrak{M}$ de un espacio afín $\mathfrak{A}$ todo conjunto linealmente independiente de puntos, cuya adherencia lineal coincide con $\mathfrak{M}$. Está claro que un conjunto de puntos $A_i (i \in I)$ de un plano $\mathfrak{M}$ es una base del plano $\mathfrak{M}$ cuando, y sólo cuando, el conjunto de vectores $\overrightarrow{A_\alpha A_i}$ $(i \in I)$, donde A_α $(\alpha \in I)$ es un punto arbitrario fijo, es una base del espacio vectorial $L(\mathfrak{M})$. En virtud del teorema sobre las bases de los espacios vectoriales del p. 4.2 tenemos:

a) *todo plano de un espacio afín posee una base;*

b) *si un plano $\mathfrak{M}$ está contenido en un plano $\mathfrak{N}$, toda base del plano $\mathfrak{M}$ es una parte de una base convenientemente escogida de $\mathfrak{N}$;*

c) *todas las bases de cualquier plano fijo $\mathfrak{M}$ son de una misma potencia igual a $\dim L(\mathfrak{M}) + 1$.*

Se llama dimensión de un plano $\mathfrak{M}$ de un espacio afín $\mathfrak{A}$ la dimensión del espacio vectorial $L(\mathfrak{M})$ que le corresponde. De la propiedad c) se deduce que la dimensión de un plano es igual a la potencia de cualquier base suya disminuida en 1. El conjunto vacío de vectores de un espacio afín $\mathfrak{A}$ se llama a veces *plano vacío* de $\mathfrak{A}$ de base vacía. La dimensión del plano vacío se toma, por definición, igual a -1.

Los planos 0-dimensionales son los puntos del espacio $\mathfrak{A}$. Los planos unidimensionales se llaman *rectas* (o líneas rectas) del espacio $\mathfrak{A}$. Cualesquiera dos puntos distintos de una recta forman una base de la misma y por cualesquiera dos puntos distintos de un espacio $\mathfrak{A}$ pasa una recta, y sólo una. En general, cualesquiera $k+1$ puntos linealmente independientes de un plano de dimensión k forman una base de este plano y por esto por cualesquiera $k+1$ puntos linealmente independientes (k es un número finito) de un espacio afín $\mathfrak{A}$ pasa un plano k-dimensional de este espacio y sólo uno.

Siendo $\mathfrak{M}$ y $\mathfrak{N}$ dos planos cualesquiera de un espacio afín $\mathfrak{A}$, convendremos en indicar por $\mathfrak{M} \vee \mathfrak{N}$ la adherencia lineal del conjunto $\mathfrak{M} \cup \mathfrak{N}$. El plano $\mathfrak{M} \vee \mathfrak{N}$ se puede definir también como la intersección de todos los planos del espacio $\mathfrak{A}$ que contienen a los planos $\mathfrak{M}$ y $\mathfrak{N}$.

La operación $\vee$, considerada como una operación binaria definida sobre el conjunto de todos los planos de un espacio afín, es conmutativa, asociativa e idempotente (es decir, $\mathfrak{M} \vee \mathfrak{M} = \mathfrak{M}$); además, para cualesquiera planos $\mathfrak{M}$ y $\mathfrak{N}$ de un espacio $\mathfrak{A}$ se tiene

$$\mathfrak{M} \subseteq \mathfrak{N} \Leftrightarrow \mathfrak{M} \vee \mathfrak{N} = \mathfrak{N}.$$

Por ejemplo, siendo $X_1, \ldots, X_m$ un conjunto finito de puntos de un espacio $\mathfrak{A}$, se tiene que

$$\mathfrak{M} = X_1 \vee X_2 \vee \ldots \vee X_m$$

es la adherencia lineal de dicho conjunto, es decir, es el menor plano que pasa por los puntos $X_1, \ldots, X_m$. La dimensión de $\mathfrak{M}$ es igual al número máximo de puntos linealmente independientes en el conjunto $X_1, \ldots, X_m$ disminuido en 1. En particular,

$$\dim. (X_1 \vee X_2 \vee \ldots \vee X_m) \leqslant m - 1.$$

Es fácil ver que tiene lugar el siguiente teorema general:

TEOREMA 2. *Cualesquiera que sean los planos $\mathfrak{M}$ y $\mathfrak{N}$ de un espacio afín $\mathfrak{A}$ se tiene*

$$\dim. (\mathfrak{M} \vee \mathfrak{N}) \leqslant \dim. \mathfrak{M} + \dim. \mathfrak{N} + 1. \tag{1}$$

Si los planos $\mathfrak{M}$ y $\mathfrak{N}$ poseen un punto común, se tiene

$$\dim. (\mathfrak{M} \vee \mathfrak{N}) + \dim. (\mathfrak{M} \cap \mathfrak{N}) = \dim. \mathfrak{M} + \dim. \mathfrak{N}. \tag{2}$$

Demostremos primero la igualdad (2). Sea $O \in \mathfrak{M} \cap \mathfrak{N}$. Entonces, la aplicación $X \longrightarrow \overrightarrow{OX}$ del espacio $\mathfrak{A}$ sobre el espacio vectorial $L(\mathfrak{A})$ transforma los planos de $\mathfrak{A}$ que pasan por el punto O en los subespacios lineales del espacio vectorial $L(\mathfrak{A})$, con la particularidad de que conserva las dimensiones y la relación de inclusión. Por esto la fórmula (2) es un corolario directo de la fórmula análoga que tiene lugar para los subespacios lineales de los espacios vectoriales (véase el teorema 1 del p. 6.1).

La estimación (1) se deduce de la fórmula (2). Efectivamente, si los planos $\mathfrak{M}$ y $\mathfrak{N}$ poseen un punto común, tenemos según (2)

$$\dim. (\mathfrak{M} \vee \mathfrak{N}) \leqslant \dim. \mathfrak{M} + \dim. \mathfrak{N}. \tag{3}$$

Sea $\mathfrak{M} \cap \mathfrak{N} = \varnothing$. Agregando el punto $O \in \mathfrak{M}$ a una base del plano $\mathfrak{N}$, obtenemos una base de $O \vee \mathfrak{N}$, es decir,

$$\dim. (O \vee \mathfrak{N}) = \dim. \mathfrak{N} + 1. \tag{4}$$

Puesto que los planos $\mathfrak{M}$ y $O \vee \mathfrak{N}$ poseen el punto común O, tenemos basándonos en (3)

$$\dim.(\mathfrak{M} \vee \mathfrak{N}) = \dim.(\mathfrak{M} \vee O \vee \mathfrak{N}) \leqslant \dim.\mathfrak{M} + \dim.(O \vee \mathfrak{N}),$$

y de aquí, debido a (4), obtenemos (1).

Por lo tanto, si los planos $\mathfrak{M}$ y $\mathfrak{N}$ no tienen puntos comunes, la dimensión de $\mathfrak{M} \vee \mathfrak{N}$ está comprendida en los límites siguientes:

$$\text{máx}(\dim.\mathfrak{M}, \dim.\mathfrak{N}) < \dim.(\mathfrak{M} \vee \mathfrak{N}) \leqslant \dim.\mathfrak{M} + \dim.\mathfrak{N} + 1.$$

Es fácil comprobar, considerando ejemplos, que la dimensión de $\mathfrak{M} \vee \mathfrak{N}$ puede tomar efectivamente cualquier valor comprendido en estos límites.

A todo plano $\mathfrak{M}$ de un espacio afín arbitrario $\mathfrak{A}$ le corresponde el subespacio lineal $L(\mathfrak{M})$ del espacio vectorial $L(\mathfrak{A})$. Según el § 6 (véase el problema 3 de la pág. 113), en $L(\mathfrak{A})$ existe *un subespacio complementario* $\mathfrak{L}_1$ que satisface las condiciones

$$L(\mathfrak{M}) + \mathfrak{L}_1 = L(\mathfrak{A}) \quad \text{y} \quad L(\mathfrak{M}) \cap \mathfrak{L}_1 = o,$$

donde o es el vector nulo del espacio $L(\mathfrak{A})$. La dimensión de $\mathfrak{L}_1$, que es la misma para todos los subespacios complementarios, se llama *codimensión* de $L(\mathfrak{M})$ en $L(\mathfrak{A})$. Por definición, se llama *codimensión de un plano* $\mathfrak{M}$ *en un espacio* $\mathfrak{A}$ la codimensión del subespacio vectorial $L(\mathfrak{M})$ en el espacio vectorial $L(\mathfrak{A})$.

Esta definición puede darse de otra forma. Para todo plano $\mathfrak{M}$ diremos que un plano $\mathfrak{N}$ es *complementario de* $\mathfrak{M}$ *en el espacio* $\mathfrak{A}$, si $\mathfrak{M} \vee \mathfrak{N} = \mathfrak{A}$ y si la intersección $\mathfrak{M} \cap \mathfrak{N}$ consta sólo de un punto. Está claro que, en estas condiciones, el subespacio $L(\mathfrak{N})$ es complementario de $L(\mathfrak{M})$ en $L(\mathfrak{A})$ y por ello *la codimensión de un plano* $\mathfrak{M}$ *en el espacio* $\mathfrak{A}$ *es igual a la dimensión de cualquier plano complementario.*

Puesto que los planos recíprocamente complementarios poseen un punto común, obtenemos, aplicándoles la fórmula (2), la igualdad

$$\dim.\mathfrak{M} + \text{codim.}\mathfrak{M} = \dim.\mathfrak{A}. \tag{5}$$

Si el espacio $\mathfrak{A}$ es de dimensión finita, obtenemos de (5)

$$\text{codim.}\mathfrak{M} = \dim.\mathfrak{A} - \dim.\mathfrak{M}. \tag{6}$$

Para un espacio $\mathfrak{A}$ de dimensión finita obtenemos de las fórmulas (2) y (6) el resultado siguiente:

TEOREMA 3. *Para cualesquiera dos planos* $\mathfrak{M}$ *y* $\mathfrak{N}$ *que se intersecan de un espacio afín* $\mathfrak{A}$ *se tiene*

$$\text{codim.}(\mathfrak{M} \cap \mathfrak{N}) + \text{codim.}(\mathfrak{M} \vee \mathfrak{N}) = \text{codim.}\mathfrak{M} + \text{codim.}\mathfrak{N}. \tag{7}$$

No obstante, es fácil ver que la fórmula (7) es válida también para los espacios de dimensión infinita.

COROLARIO. *Si la intersección de los planos* $\mathfrak{N}_1, \mathfrak{N}_2, \ldots, \mathfrak{N}_s$ *de un espacio afín* $\mathfrak{A}$ *arbitrario es no vacía, se tiene*

$$\text{codim.}\,(\mathfrak{N}_1 \cap \ldots \cap \mathfrak{N}_s) \leqslant \text{codim.}\,\mathfrak{N}_1 + \ldots + \text{codim.}\,\mathfrak{N}_s. \qquad (8)$$

Para $s=2$, la fórmula (8) se desprende directamente de (7) y en el caso general ella se obtiene aplicando sucesivamente la fórmula del caso particular indicado.

Si el plano $\mathfrak{M}$ se encuentra sobre el plano $\mathfrak{N}$ de un espacio $\mathfrak{A}$, entonces, considerando $\mathfrak{N}$ como el espacio afín sobre $L(\mathfrak{N})$ e indicando por $\text{codim.}_{\mathfrak{N}}\mathfrak{M}$ la codimensión del plano $\mathfrak{M}$ en el espacio $\mathfrak{N}$, obtenemos la fórmula

$$\text{codim.}_{\mathfrak{N}}\mathfrak{M} + \text{codim.}\,\mathfrak{N} = \text{codim.}\,\mathfrak{M} \qquad (9)$$

que se demuestra fácilmente pasando a los subespacios del espacio vectorial $L(\mathfrak{A})$. De la fórmula (9) se deduce, en particular, que si la codimensión del plano $\mathfrak{M}$ en $\mathfrak{A}$ es finita y $\mathfrak{M} \subset \mathfrak{N}$, se tiene

$$\text{codim.}\,\mathfrak{N} \leqslant \text{codim.}\,\mathfrak{M} - 1. \qquad (10)$$

Los planos de codimensión igual a 1 se llaman *hiperplanos*. En otras palabras, en un espacio afín de dimensión finita n se llaman hiperplanos los planos de dimensión $n-1$. Luego, por cada n puntos linealmente independientes de un espacio de este tipo pasa un hiperplano, y sólo uno.

Consideremos ahora un conjunto cualquiera finito de hiperplanos $\mathfrak{P}_1, \ldots, \mathfrak{P}_s$ de un espacio afín arbitrario $\mathfrak{A}$. Debido a la fórmula (8) tenemos

$$0 < \text{codim.}\,(\mathfrak{P}_1 \cap \ldots \cap \mathfrak{P}_s) \leqslant s. \qquad (11)$$

Se dice que los hiperplanos $\mathfrak{P}_1, \ldots, \mathfrak{P}_s$ (para s no mayor que la dimensión de $\mathfrak{A}$) se encuentran *en posición general*, si tiene lugar la igualdad exacta

$$\text{codim.}\,(\mathfrak{P}_1 \cap \ldots \cap \mathfrak{P}_s) = s. \qquad (12)$$

Probemos que para todo $t \leqslant s$ de (12) resulta

$$\text{codim.}\,(\mathfrak{P}_1 \cap \ldots \cap \mathfrak{P}_t) = t.$$

En efecto, si fuese

$$\text{codim.}\,(\mathfrak{P}_1 \cap \ldots \cap \mathfrak{P}_t) < t,$$

tendríamos, debido a (8) y a (11),

$$\text{codim.}\,((\mathfrak{P}_1 \cap \ldots \cap \mathfrak{P}_t) \cap (\mathfrak{P}_{t+1} \cap \ldots \cap \mathfrak{P}_s)) < t + (s-t) = s$$

lo que estaría en contradicción con (12).

TEOREMA 4. *Todo plano* $\mathfrak{M}$ *de una codimensión finita* s *es una intersección de* s *hiperplanos convenientemente escogidos.*

Tomemos una base de $\mathfrak{M}$ y complementémosla con unos puntos $A_1, \ldots, A_s$ hasta obtener una base de todo el espacio $\mathfrak{A}$. Sea $\mathfrak{P}_i$ la adherencia lineal del conjunto $\mathfrak{M} \cup \{A_1, \ldots, A_{i-1}, A_{i+1}, \ldots, A_s\}$. Como $\mathfrak{P}_i \neq \mathfrak{A}$ y $\mathfrak{P}_i \vee A_i = \mathfrak{A}$, resulta que $\mathfrak{P}_i$ es un hiperplano ($i = 1, \ldots, s$) y que

$$\mathfrak{M} \subseteq \mathfrak{P}_1 \cap \ldots \cap \mathfrak{P}_s.$$

De $A_i \notin \mathfrak{P}_i$ tenemos

$$\mathfrak{A} \supset \mathfrak{P}_1 \supset \mathfrak{P}_1 \cap \mathfrak{P}_2 \supset \ldots \supset \mathfrak{P}_1 \cap \ldots \mathfrak{P}_s,$$

de donde, debido a (10),

$$\text{codim.}\,(\mathfrak{P}_1 \cap \ldots \cap \mathfrak{P}_s) \geqslant s.$$

Comparando esta desigualdad con (11), obtenemos

$$\text{codim.}\,(\mathfrak{P}_1 \cap \ldots \cap \mathfrak{P}_s) = s.$$

Puesto que las codimensiones de los planos $\mathfrak{M}$ y $\mathfrak{P}_1 \cap \ldots \cap \mathfrak{P}_s$ coinciden y el primero está contenido en el segundo, ambos planos, debido a (11), coinciden que es lo que se quería demostrar.

Para terminar, haremos unas observaciones relacionadas con la definición del concepto del plano aceptada más arriba. Esta definición puede ser enunciada de nuevo en la forma siguiente: un conjunto $\mathfrak{M}$ de puntos de un espacio afín $\mathfrak{A}$ se llama plano, si para todo número natural s y para toda sucesión $A_1, \ldots, A_s$ de puntos de $\mathfrak{M}$ cualquier punto del espacio $\mathfrak{A}$, que dependa linealmente de los puntos de la sucesión mencionada, pertenece a $\mathfrak{M}$. De hecho, es suficiente tomar en esta definición $s = 3$ y para varios tipos de cuerpos conmutativos principales K es suficiente tomar incluso $s = 2$. Tiene lugar concretamente el teorema siguiente:

TEOREMA 5 *Si un conjunto $\mathfrak{M}$ de puntos de un espacio afín $\mathfrak{A}$ contiene con tres cualesquiera puntos suyos todos los puntos de $\mathfrak{A}$ que dependen linealmente de éstos, el conjunto $\mathfrak{M}$ es un plano. Si $\mathfrak{A}$ es un espacio afín sobre un cuerpo conmutativo K, de característica diferente de 2, y si un conjunto $\mathfrak{M}$ con dos cualesquiera puntos suyos contiene todos los puntos de $\mathfrak{A}$ que dependen linealmente de éstos, el conjunto $\mathfrak{M}$ es un plano de $\mathfrak{A}$.*

Demostremos la segunda afirmación. Supongamos que un conjunto $\mathfrak{M}$ cumple las condiciones de esta afirmación y que $X_0, X_1, X_2 \ldots$ es una sucesión de puntos de $\mathfrak{M}$. Debemos demostrar que para cualesquiera $\lambda_1, \ldots, \lambda_s \in K$ la relación

$$\overrightarrow{X_0X} = \lambda_1\overrightarrow{X_0X_1} + \ldots + \lambda_s\overrightarrow{X_0X_s}$$

implica $X \in \mathfrak{M}$. Para $s = 1$ esto es válido por el enunciado mismo del teorema. Aplicando ahora la inducción, aceptamos que la implicación indicada es válida para un $s \geqslant 1$. Sea

$$\overrightarrow{X_0Y} = \lambda_1\overrightarrow{X_0X_1} + \ldots + \lambda_{s+1}\overrightarrow{X_0X_{s+1}}.$$

Consideremos unos puntos X y Z que satisfacen las condiciones

$$\overrightarrow{X_0X} = \lambda_1\overrightarrow{X_0X_1} + \ldots + \lambda_s\overrightarrow{X_0X_s} \text{ y } \overrightarrow{X_0Z} = \lambda_{s+1}\overrightarrow{X_0X_{s+1}}.$$

Según la hipótesis de inducción, de estas relaciones resulta que $X \in \mathfrak{M}$ y $Z \in \mathfrak{M}$. Tomando ahora (fig. 5)

$$\overrightarrow{X_0B} = \frac{1}{2}(\overrightarrow{X_0X} + \overrightarrow{X_0Z}),$$

vemos que el punto B depende linealmente de X y de Z y que el punto Y depende linealmente de los puntos X_0 y B. Basándonos en las condiciones del teorema, obtenemos de aquí sucesivamente que $B \in \mathfrak{M}$ e $Y \in \mathfrak{M}$ que es lo que se quería demostrar. La primera afirmación del teorema 5 se demuestra análogamente, pero sin emplear el punto B, y por esto ella es válida independientemente de la característica del cuerpo conmutativo K. Es fácil ver que para cuerpos conmutativos de característica 2 la segunda afirmación no tiene lugar.

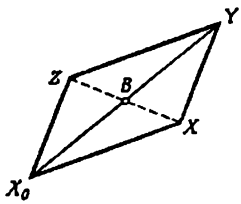


Fig. 5.

Hemos introducido en el p. 29.1 dos conceptos principales: el de un espacio afín sobre un espacio vectorial y el de un espacio afín sobre un cuerpo conmutativo. En este punto hemos considerado hasta el momento los espacios afines sobre los espacios vectoriales. Sea ahora $\mathfrak{A}$ un espacio afín sobre un cuerpo conmutativo K y sean $S(X, Y, Z)$ y $P_\lambda(X, Y)$ las operaciones principales de $\mathfrak{A}$. Hemos visto ya en el p. 29.1 que las traslaciones de $\mathfrak{A}$ son unos automorfismos del espacio $\mathfrak{A}$ sobre K de tipo especial. Demostremos ahora que *un punto X de un espacio $\mathfrak{A}$ depende linealmente de unos puntos $X_0, X_1, \ldots X_m$ de $\mathfrak{A}$ cuando, y sólo cuando, X se expresa en forma de un polinomio en $X_0, \ldots X_m$ mediante las operaciones S y P_λ*.

En efecto, si

$$\overrightarrow{X_0X} = \lambda_1\overrightarrow{X_0X_1} + \ldots + \lambda_m\overrightarrow{X_0X_m},$$

donde $\lambda_1, \ldots, \lambda_m \in K$, se tiene según el p. 29.1

$$X = ((X_0 \cdot \lambda_1\overrightarrow{X_0X_1}) \ldots) \cdot \lambda_m\overrightarrow{X_0X_m}. \tag{13}$$

Pero para cualesquiera $B, C, Z \in \mathfrak{A}$ y $\lambda \in K$ se tiene

$$Z \cdot \lambda\overrightarrow{BC} = S(B, P_\lambda(B, C), Z).$$

Transformando mediante esta fórmula el segundo miembro de la igualdad (13), obtenemos la expresión requerida de X en forma de un polinomio en $X_0, X_1, \ldots, X_m$. Recíprocamente, de

$$X = S(Y_1, Y_2, Y_3) \text{ y } Z = P_\lambda(Y_1, Y_2)$$

se deduce que

$$\overline{Y_3X} = \overline{Y_1Y_3} \quad \text{e} \quad \overline{Y_1Z} = \lambda\overline{Y_1Y_2},$$

es decir, que X y Z dependen linealmente de Y_1, Y_2, Y_3 y por esto, si X se expresa en forma de un polinomio en $X_0, X_1, \ldots, X_m$, el punto X depende linealmente de $X_0, X_1, \ldots, X_m$.

Según la definición, un conjunto $\mathfrak{M}$ de puntos de un espacio $\mathfrak{A}$ se llama plano, si este conjunto $\mathfrak{M}$ con cualesquiera puntos suyos $X_1, \ldots, X_m$ contiene también todos los puntos de $\mathfrak{A}$ que dependen linealmente de éstos, es decir, contiene los valores de todos los polinomios en $X_1, \ldots, X_m$. En otras palabras, los planos de un espacio $\mathfrak{A}$ son simplemente subálgebras del álgebra $\mathfrak{A}$ y las adherencias lineales son las subálgebras generadas por los elementos de los conjuntos correspondientes.

29.3. Planos paralelos. Hemos visto ya que para todo plano $\mathfrak{M}$ de un espacio afín $\mathfrak{A}$ el conjunto $L(\mathfrak{M})$ de todos los vectores de tipo $\overline{XY}$ $(X, Y \in \mathfrak{M})$ es un subespacio del espacio vectorial $L(\mathfrak{A})$. El subespacio $L(\mathfrak{M})$ suele llamarse frecuentemente *subespacio tangente* al plano $\mathfrak{M}$.

DEFINICIÓN. *Los planos $\mathfrak{M}$ y $\mathfrak{N}$ de un espacio $\mathfrak{A}$ se llaman paralelos (notación simbólica $\mathfrak{M} \parallel \mathfrak{N}$), si sus espacios vectoriales tangentes $L(\mathfrak{M})$ y $L(\mathfrak{N})$ son incidentes, es decir, si $L(\mathfrak{M}) \subseteq L(\mathfrak{N})$ o $L(\mathfrak{N}) \subseteq L(\mathfrak{M})$.*

De esta definición se desprenden inmediatamente varios corolarios importantes.

COROLARIO 1. *Dos planos $\mathfrak{M}_1$ y $\mathfrak{M}_2$ que se intersecan son paralelos cuando, y sólo cuando, uno de ellos está contenido en el otro.*

Efectivamente, si $\mathfrak{M}_1$ y $\mathfrak{M}_2$ tienen un punto común O, entonces $L(\mathfrak{M}_i)$ $(i = 1, 2)$ es el conjunto de vectores de tipo $\overline{OX}$ $(X \in \mathfrak{M}_i)$ y por ello la incidencia de los espacios tangentes $L(\mathfrak{M}_1)$ y $L(\mathfrak{M}_2)$ equivale a la incidencia de los planos $\mathfrak{M}_1$ y $\mathfrak{M}_2$.

COROLARIO 2. *Si los planos $\mathfrak{M}_1$, $\mathfrak{M}_2$ y $\mathfrak{M}_3$ son de una misma dimensión finita (o de una misma codimensión finita) y $\mathfrak{M}_1 \parallel \mathfrak{M}_3$ y $\mathfrak{M}_2 \parallel \mathfrak{M}_3$, se tiene $\mathfrak{M}_1 \parallel \mathfrak{M}_2$.*

Las dimensiones (las codimensiones) de los espacios tangentes $L(\mathfrak{M}_i)$ son finitas e iguales. Por ello la incidencia de $L(\mathfrak{M}_i)$ y $L(\mathfrak{M}_j)$ equivale a la coincidencia de los mismos y la afirmación del corolario se reduce a la observación trivial de que $L(\mathfrak{M}_1) = L(\mathfrak{M}_3)$ y $L(\mathfrak{M}_2) = L(\mathfrak{M}_3)$ implican $L(\mathfrak{M}_1) = L(\mathfrak{M}_2)$.

El corolario 2 significa que para cualquier k finito todos los planos k-dimensionales de un espacio afín arbitrario $\mathfrak{A}$ se descomponen en haces de planos paralelos entre sí. Estos haces se llaman a veces direcciones *k-dimensionales en $\mathfrak{A}$*.

COROLARIO 3. *En un espacio afín $\mathfrak{A}$ se tiene*

$$\mathfrak{M} \parallel \mathfrak{M} \cdot u$$

para cualquier plano $\mathfrak{M}$ y para cualquier traslación $u \in L(\mathfrak{A})$.

Efectivamente, el espacio tangente $L(\mathfrak{M}\cdot u)$ está formado por vectores de tipo $\overline{XuYu}$ (X, $Y\in\mathfrak{M}$). Pero

$$\overline{XuYu}=\overline{XuX}+\overline{XY}+\overline{YYu}=-u+\overline{XY}+u=\overline{XY}$$

y por ello $L(\mathfrak{M}\cdot u)=L(\mathfrak{M})$.

COROLARIO 4. *Por todo punto B de un espacio afín pasa un hiperplano $\mathfrak{P}$, y sólo uno, paralelo a cualquier hiperplano $\mathfrak{M}$ dado de antemano.*

Si $B\in\mathfrak{M}$, se ve del corolario 1 que el único hiperplano que pasa por B y es paralelo a $\mathfrak{M}$, es el propio hiperplano $\mathfrak{M}$. Podemos aceptar, pues, que $B\notin\mathfrak{M}$. En virtud del corolario 3, el hiperplano $\mathfrak{M}\cdot\overline{OB}$ ($O\in\mathfrak{M}$) pasa por B y es paralelo a $\mathfrak{M}$. Si existe otro hiperplano $\mathfrak{N}$ que pasa por B y tal que $\mathfrak{N}\parallel\mathfrak{M}$, tenemos, según el corolario 2, $\mathfrak{N}\parallel\mathfrak{M}\cdot\overline{OB}$ y entonces, en virtud del corolario 1, $\mathfrak{N}$ y $\mathfrak{M}\cdot\overline{OB}$ son incidentes. Pero los hiperplanos incidentes coinciden que es lo que se quería demostrar.

TEOREMA 1. *Si un hiperplano $\mathfrak{P}$ y un plano $\mathfrak{N}$ de un espacio afín $\mathfrak{A}$ no se intersecan, resulta que $\mathfrak{P}$ es paralelo a $\mathfrak{N}$.*

Queremos demostrar que $L(\mathfrak{N})\subseteq L(\mathfrak{P})$. Supongamos, al contrario, que para unos puntos A, $B\in\mathfrak{N}$ el vector $\overline{AB}$ no pertenece a $L(\mathfrak{P})$. Tomemos un punto cualquiera $O\in\mathfrak{P}$ (fig. 6) y consideremos el vector $\overline{OA}$. Puesto que $\mathfrak{P}$ es un hiperplano, se tiene

$$L(\mathfrak{A})=L(\mathfrak{P})+L(A\vee B)$$

y por ello para $X\in\mathfrak{P}$ y $U\in A\vee B$ convenientemente escogidos tenemos $\overline{OA}=\overline{OX}-\overline{AU}$, de donde $\overline{OU}=\overline{OX}$ y, por consiguiente, $U=X$, lo que contradice a la condición de que $\mathfrak{P}\cap\mathfrak{N}=\varnothing$.

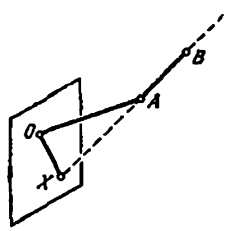


Fig. 6.

TEOREMA 2. *Para que dos planos $\mathfrak{M}$ y $\mathfrak{N}$ que no se intersecan sean paralelos, es necesario y suficiente que uno de ellos sea un hiperplano en el espacio $\mathfrak{M}\vee\mathfrak{N}$.*

La suficiencia está contenida directamente en el teorema 1. Demostremos la necesidad. Sea $L(\mathfrak{N})\subseteq L(\mathfrak{M})$ y sea $A\in\mathfrak{N}$. Es suficiente demostrar que $\mathfrak{M}\vee\mathfrak{N}\subseteq\mathfrak{M}\vee A$. El plano $\mathfrak{M}\vee\mathfrak{N}$ está formado por aquellos puntos X que dependen linealmente de unos puntos A, X_1, ..., X_s de $\mathfrak{N}$ y unos puntos Y_1, ..., Y_t de $\mathfrak{M}$. Es decir,

$$\overline{AX}=\lambda_1\overline{AY}_1+\ldots+\lambda_t\overline{AY}_t+\mu_1\overline{AX}_1+\ldots+\mu_s\overline{AX}_s=$$
$$=\overline{AY}+(\mu_1+\ldots+\mu_s)\overline{AO}+\overline{OZ}\quad(Y\in\mathfrak{N}\text{ y }O,\ Z\in\mathfrak{M}).$$

Tenemos, por hipótesis, $\overline{AY}\in L(\mathfrak{N})\subseteq L(\mathfrak{M})$ y, por consiguiente, $\overline{AY}=\overline{OU}$ ($U\in\mathfrak{M}$). De aquí se desprende que $\overline{AX}\in L(\mathfrak{M}\vee A)$ y por lo tanto $X\in\mathfrak{M}\vee A$.

COROLARIO 1. *Si los planos* $\mathfrak{M}$ *y* $\mathfrak{N}$ *son de dimensión finita y no se intersecan, estos planos son paralelos cuando, y sólo cuando,*

$$\dim.(\mathfrak{M} \vee \mathfrak{N}) = \max(\dim.\mathfrak{M},\ \dim.\mathfrak{N}) + 1. \qquad (1)$$

La demostración es obvia. En particular, de este corolario sacamos la conclusión de que dos rectas que no se intersecan son paralelas cuando, y sólo cuando, se encuentran sobre un mismo plano bidimensional. Análogamente, una recta y un plano bidimensional que no se intersecan son paralelos cuando, y sólo cuando, se encuentran en un mismo plano tridimensional, etc.

COROLARIO 2. *Cualquiera que sea el número finito* $k > 0$, *por todo punto* B *de un espacio afín* $\mathfrak{A}$ *pasa un plano* $\mathfrak{P}$ *k-dimensional, y sólo uno, paralelo a un plano k-dimensional* $\mathfrak{M}$ *dado de antemano.*

Podemos limitarnos a considerar el caso en que $B \notin \mathfrak{M}$. Debido a la fórmula (1), el plano $\mathfrak{P}$ debe ser un hiperplano del espacio $\mathfrak{M} \vee B$ y por ello el asunto se reduce al corolario 4 de la definición de los planos paralelos.

29.4. Funcionales lineales. Una función $f: \mathfrak{A} \to K$, definida sobre el conjunto de todos los puntos de un espacio afín $\mathfrak{A}$ y con valores en el cuerpo conmutativo principal K, se llama *funcional* sobre $\mathfrak{A}$.

El concepto de una funcional sobre $\mathfrak{A}$ es un caso particular del concepto de una función que está definida sobre un conjunto arbitrario y cuyos valores pertenecen a un cuerpo conmutativo dado K. Para estas funciones hemos definido en el p. 4.1 los conceptos de una suma y de un producto de un número por una función, resultando que el conjunto de todas las funciones forma, respecto a estas operaciones, un espacio lineal, cuya dimensión coincide con la potencia del conjunto que sirve como el dominio de definición.

Una funcional f, definida sobre un espacio afín $\mathfrak{A}$, se llama *lineal* (o *afín*) sobre $\mathfrak{A}$, si para cualesquiera $O, P, Q, R \in \mathfrak{A}$ y $\lambda, \mu \in K$ que cumplen la relación

$$\overrightarrow{OR} = \lambda \overrightarrow{OP} + \mu \overrightarrow{OQ} \qquad (1)$$

tiene lugar la igualdad

$$f(R) = \lambda(f(P) - f(O)) + \mu(f(Q) - f(O)) + f(O). \qquad (2)$$

Demostremos que para toda funcional lineal f es válida la implicación

$$\overrightarrow{X_0X} = \sum_{i=1}^{s} \lambda_i \overrightarrow{X_0X_i} \Rightarrow f(X) = \sum_{i=1}^{s} \lambda_i (f(X_i) - f(X_0)) + f(X_0), \qquad (3)$$

cualesquiera que sean $X_0, X_1, \ldots, X_s, X \in \mathfrak{A}$ y $\lambda_1, \ldots, \lambda_s \in K$.

Para $s = 2$ la implicación (3) coincide la implicación (1) $\Rightarrow$ (2) que define la linealidad de f. Aceptemos ahora por inducción que la implicación (3) es válida para un valor fijo $s \geqslant 2$. Sea $\overrightarrow{X_0Y} = \sum_{i=1}^{s+1} \lambda_i \overrightarrow{X_0X_i}$. Con-

siderando un punto X que cumple la condición $\overrightarrow{X_0X} = \sum_{i=1}^{s} \lambda_i \overrightarrow{X_0X_i}$, obtenemos (3). Después, de la relación $\overrightarrow{X_0Y} = \overrightarrow{X_0X} + \lambda_{s+1} \overrightarrow{X_0X_{s+1}}$ y de la implicación (1) $\Rightarrow$ (2) obtenemos

$$f(Y) = (f(X) - f(X_0)) + \lambda_{s+1}(f(X_{s+1}) - f(X_0)) + f(X)_0 =$$
$$= \sum_{i=1}^{s+1} \lambda_i (f(X_i) - f(X_0)) + f(X_0)$$

que es lo que se quería demostrar.

Está claro que todas las funcionales constantes son lineales. Es también evidente que la suma de unas funcionales lineales y el producto de un número por una funcional lineal son de nuevo funcionales lineales. Por esto el conjunto de todas las funcionales lineales sobre un espacio afín $\mathfrak{A}$ es un espacio lineal. Este espacio se indica por $\mathfrak{A}_f$ y se llama *conjugado* de $\mathfrak{A}$. El conjunto de todas las funcionales constantes es un subespacio lineal de dimensión uno del espacio $\mathfrak{A}_f$.

LEMA. *Si f es una funcional lineal sobre un espacio afín $\mathfrak{A}$ y para unos puntos A, $B \in \mathfrak{A}$ se tiene $f(A) \neq f(B)$, entonces en la recta $A \vee B$ existe un punto X tal que $f(X) = \alpha$ cualquiera que sea $\alpha \in K$.*

Tenemos para un punto arbitrario $X \in A \vee B$

$$\overrightarrow{AX} = \lambda \cdot \overrightarrow{AB} \quad (\lambda \in K), \tag{4}$$

de donde, en virtud de (2),

$$f(X) = \lambda (f(B) - f(A)) + f(A).$$

Por ello, tomando para λ el valor

$$\lambda = (\alpha - f(A))(f(B) - (f(A))^{-1},$$

obtenemos de (4) el punto requerido $X \in A \vee B$ tal que $f(X) = \alpha$.

TEOREMA 1. *Para toda funcional lineal no constante f sobre un espacio afín $\mathfrak{A}$, el conjunto de los puntos $X \in \mathfrak{A}$ en los que f toma un valor fijo $\alpha \in K$ es un hiperplano de $\mathfrak{A}$. Recíprocamente, para todo hiperplano $\mathfrak{M}$ y cualquier $\alpha \in K$ existe una funcional lineal f, cuyos valores son iguales a α sobre $\mathfrak{M}$ y diferentes de α fuera de $\mathfrak{M}$.*

Demostremos la primera afirmación. Sea $\mathfrak{M}$ el conjunto de los puntos $X \in \mathfrak{A}$ en los que $f(X) = \alpha$. Puesto que la funcional f no es constante sobre $\mathfrak{A}$, tendremos $f(A) \neq f(B)$ para unos puntos A, $B \in \mathfrak{A}$ convenientemente escogidos. Según el lema, de aquí se deduce que en la recta $A \vee B$ existe un punto X perteneciente a $\mathfrak{M}$, de modo que el conjunto $\mathfrak{M}$ no es vacío. Si $X_0, X_1, \ldots, X_s \in \mathfrak{M}$ y X es un punto que depende linealmente de éstos, se tiene

$$f(X_0) = f(X_1) = \ldots = f(X_s) = \alpha \tag{5}$$

y para unos valores adecuados $\lambda_1, \ldots, \lambda_s \in K$ se tiene

$$\overrightarrow{X_0X} = \lambda_1 \cdot \overrightarrow{X_0X_1} + \ldots + \lambda_s \cdot \overrightarrow{X_0X_s}. \tag{6}$$

De las fórmulas (6), (3) y (5) encontramos $f(X) = \alpha$, es decir, $X \in \mathfrak{M}$. En otras palabras, el conjunto $\mathfrak{M}$ es *un plano* de $\mathfrak{A}$. Resta demostrar que codim. $\mathfrak{M} = 1$. Como la funcional f no es constante, existe un punto A tal que $f(A) \neq \alpha$ y por ello $A \notin \mathfrak{M}$. Tomemos un punto cualquiera $X \in \mathfrak{A}$ y demostremos que $X \in A \vee \mathfrak{M}$. Si $f(X) = \alpha$, tenemos $X \in \mathfrak{M}$ y no hay nada que demostrar. Sea, por esto, $f(X) \neq \alpha$ y sea $M \in \mathfrak{M}$. En la recta $A \vee M$ (fig. 7) existe, según el lema, un punto B en el cual $f(B) \neq f(X)$ y, por ello, en la recta $B \vee X$ existe un punto $N \in \mathfrak{M}$. Vemos que X depende linealmente de B y de N y que B depende linealmente de A y M. Luego, X depende linealmente de A, M y N y $X \in \mathfrak{M} \vee A$.

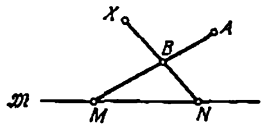


Fig. 7.

Demostremos ahora la segunda afirmación del teorema. Sea $\mathfrak{A} = \mathfrak{M} \vee A$ y sea $M \in \mathfrak{M}$. Para todo punto $X \in \mathfrak{A}$ existe entonces una descomposición de tipo

$$\overrightarrow{MX} = \lambda \cdot \overrightarrow{MA} + \overrightarrow{MU} \quad (\lambda \in K \text{ y } U \in \mathfrak{M})$$

y sólo una. Introduciendo la funcional $f_0(X) = \lambda$, vemos que ella es igual a 0 sobre $\mathfrak{M}$ y es diferente de 0 fuera de $\mathfrak{M}$ y que la funcional $f(X) = f_0(X) + \alpha$ es igual a α sobre $\mathfrak{M}$ y es diferente de α fuera de $\mathfrak{M}$. Por lo tanto, resta sólo demostrar que la funcional f_0 es lineal. Sea

$$\begin{aligned} \overrightarrow{X_0X_{s+1}} &= \lambda_1 \cdot \overrightarrow{X_0X_1} + \ldots + \lambda_s \cdot \overrightarrow{X_0X_s}, \\ \overrightarrow{MX_i} &= \alpha_i \cdot \overrightarrow{MA} + \overrightarrow{MU_i} \quad (U_i \in \mathfrak{M};\ i = 0, 1, \ldots, s) \end{aligned} \tag{7}$$

y, por consiguiente, $f_0(X_i) = \alpha_i$ $(i = 0, 1, \ldots, s)$. Como $\overrightarrow{X_0X_j} = \overrightarrow{MX_j} - \overrightarrow{MX_0}$, de las fórmulas (7) obtenemos

$$\overrightarrow{MX_{s+1}} = MX_0 + \sum \lambda_i (\overrightarrow{MX_i} - \overrightarrow{MX_0}) = (\sum (\lambda_i \alpha_i - \lambda_i \alpha_0) + \alpha_0) \cdot \overrightarrow{MA} + \overrightarrow{MU},$$

de donde

$$f_0(X_{s+1}) = \sum \lambda_i (f(X_i) - f(X_0)) + f(X_0)$$

que es lo que se quería demostrar.

El conjunto de aquellos puntos $X \in \mathfrak{A}$ en los que la funcional dada f o el sistema de funcionales f_i dado se anula se llama *variedad radical* de la funcional (o del sistema de funcionales). Por esto el teorema 1 equivale a la afirmación de que la variedad radical de una funcional lineal no constante es un hiperplano y de que todo hiperplano es la variedad radical de una funcional lineal ade-

cuada. Si la funcional es constante, su variedad radical es, obviamente, o bien vacía o bien coincidente con $\mathfrak{A}$ (funcional nula). La variedad radical de un sistema de funcionales es la intersección de las variedades radicales correspondientes a las funcionales del sistema dado. Por lo tanto, la variedad radical de un sistema arbitrario de funcionales lineales es un plano (posiblemente vacío o coincidente con todo el espacio $\mathfrak{A}$).

Hemos demostrado en el p. 29.2 que todo plano $\mathfrak{M}$ de codimensión finita $s \geqslant 1$ puede ser representado en la forma $\mathfrak{M} = \mathfrak{P}_1 \cap \ldots \cap \mathfrak{P}_s$, donde $\mathfrak{P}_1, \ldots, \mathfrak{P}_s$ son unos hiperplanos adecuados. Indicando por f_i la funcional lineal, cuya variedad radical es $\mathfrak{P}_i$, vemos que $\mathfrak{M}$ es el conjunto de aquellos puntos $X \in \mathfrak{A}$, para los cuales

$$f_1(X) = 0, \ \ldots, \ f_s(X) = 0. \tag{8}$$

Considerando el problema recíproco, preguntémonos ¿para qué funcionales lineales $f_1, \ldots, f_s$ la correspondiente variedad radical (8) resultará ser un plano de codimensión s?

TEOREMA 2. *Para que la variedad radical de un sistema de funcionales lineales $f_1, \ldots, f_s$ sea un plano de codimensión s, es necesario y suficiente que las funcionales $f_1, \ldots, f_s$ sean linealmente independientes (en $\mathfrak{A}_f$) y que posean al menos un punto radical común.*

NECESIDAD. Indiquemos por $\mathfrak{P}_i$ la variedad radical de la funcional f_i. Si

$$f_s(X) = \alpha_1 f_1(X) + \ldots + \alpha_{s-1} f_{s-1}(X) \quad (X \in \mathfrak{A}),$$

se tiene $\mathfrak{P}_1 \cap \ldots \cap \mathfrak{P}_{s-1} \subseteq \mathfrak{P}_s$. Luego, resulta que codim. $(\mathfrak{P}_1 \cap \ldots \cap \mathfrak{P}_s) =$ $=$ codim. $(\mathfrak{P}_1 \cap \ldots \cap \mathfrak{P}_{s-1}) \neq s$ y, por consiguiente, las condiciones del teorema son necesarias.

SUFICIENCIA. Tomemos

$$\mathfrak{M} = \mathfrak{P}_1 \cap \ldots \cap \mathfrak{P}_{s-1} \cap \mathfrak{P}_s.$$

Tenemos, según el p. 29.2, codim. $\mathfrak{M} \leqslant s$. Supongamos que codim. $\mathfrak{M} < s$. Existen entonces en $\mathfrak{A}$ unos puntos $A_1, \ldots, A_{s-1}$ tales que

$$\mathfrak{A} = \mathfrak{M} \vee A_1 \vee \ldots \vee A_{s-1}. \tag{9}$$

Consideremos todas las funciones definidas sobre el conjunto $\{A_1, \ldots, A_{s-1}\}$ con valores en un cuerpo conmutativo K. Estas funciones forman un espacio lineal, cuya dimensión es igual a $s-1$, si los puntos $A_1, \ldots, A_{s-1}$ son diferentes, y es menor que $s-1$, si algunos de estos puntos coinciden. Es decir, el espacio de las funciones definidas sobre $\{A_1, \ldots, A_{s-1}\}$ es de dimensión no mayor que $s-1$ y por ello al menos una de s cualesquiera funciones definidas sobre el conjunto señalado puede ser expresada linealmente

en términos de las demás. En particular, si consideramos las funcionales $f_1, \ldots, f_s$ sólo sobre el conjunto $\{A_1, \ldots, A_{s-1}\}$, una de ellas puede ser sin duda alguna expresada linealmente en términos de las restantes. Sea, por ejemplo,

$$f_s(X) = \alpha_1 f_1(X) + \ldots + \alpha_{s-1} f_{s-1}(X). \tag{10}$$

Esta igualdad es válida para todos los puntos $X = A_1, \ldots, A_{s-1}$. Basándonos en que las funcionales consideradas son lineales y se anulan sobre el plano $\mathfrak{M}$ y empleando la relación (9), veremos fácilmente que la igualdad (10) es válida para todos los puntos $X \in \mathfrak{A}$, es decir, que las funcionales $f_1, \ldots, f_s$ son linealmente dependientes y con ello queda demostrada la suficiencia de las condiciones. En efecto, de (9) se deduce que para todo punto $X \in \mathfrak{A}$ existen unos puntos $O, Y \in \mathfrak{M}$ y unos números $\lambda_1, \ldots, \lambda_{s-1}$ tales que

$$\overrightarrow{OX} = \lambda_1 \overrightarrow{OA_1} + \ldots + \lambda_{s-1} \overrightarrow{OA_{s-1}} + \overrightarrow{OY};$$

luego, $f_i(O) = f_i(Y) = 0$ y en virtud e (3) tenemos

$$f_i(X) = \lambda_1 f_i(A_1) + \ldots + \lambda_{s-1} f_i(A_{s-1}) \quad (i = 1, \ldots, s). \tag{11}$$

Teniendo en cuenta que la igualdad (10) es válida para $X = A_j$ $(j = 1, \ldots, s-1)$, obtenemos de (11)

$$f_s(X) = \sum_{j=1}^{s-1} \lambda_j f_s(A_j) = \sum_{j=1}^{s-1} \lambda_j \sum_{k=1}^{s-1} \alpha_k f_k(A_j) =$$

$$= \sum_{k=1}^{s-1} \alpha_k \sum_{j=1}^{s-1} \lambda_j f_k(A_j) = \alpha_1 f_1(X) + \ldots + \alpha_{s-1} f_{s-1}(X)$$

que es lo que se quería demostrar.

COROLARIO 1. *Supongamos que un plano $\mathfrak{M}$ de un espacio afín $\mathfrak{A}$ es de codimensión finita s. Entonces el conjunto $\mathfrak{M}^\perp$ de todas las funcionales lineales sobre $\mathfrak{A}$ iguales a 0 sobre $\mathfrak{M}$, es un espacio vectorial de dimensión s.*

Efectivamente, si $\mathfrak{M}^\perp$ contiene unas funcionales $f_1, \ldots, f_{s+1}$ linealmente independientes y si $\mathfrak{P}_i$ es el hiperplano radical de la funcional f_i, tenemos

$$\mathfrak{M} \subseteq \mathfrak{P}_1 \cap \ldots \cap \mathfrak{P}_{s+1}$$

y, por ello,

$$\text{codim. } \mathfrak{M} \geqslant s + 1.$$

En cambio, si todas las funcionales lineales de $\mathfrak{M}^\perp$ se expresan linealmente en términos de $f_1, \ldots, f_{s-1}$, se tiene

$$\mathfrak{M} = \mathfrak{P}_1 \cap \ldots \cap \mathfrak{P}_{s-1}$$

y, por consiguiente, codim. $\mathfrak{M} \leqslant s - 1$.

COROLARIO 2. *Para todo espacio afín $\mathfrak{A}$ la dimensión del espacio conjugado $\mathfrak{A}_f$ es mayor en 1 que la dimensión de $\mathfrak{A}$.*

Sea $\mathfrak{B} = \{A_l \mid l \in I\}$ una base de $\mathfrak{A}$ y sea F una función cualquiera que está definida sobre $\mathfrak{B}$ y toma valores en K. Fijemos un punto cualquiera $O \in \mathfrak{B}$. Para todo punto $X \in \mathfrak{A}$ existe un sistema $\lambda_{l_1}, \ldots, \lambda_{l_s} \in K$ definido unívocamente que cumple la condición

$$\overrightarrow{OX} = \lambda_{l_1}\overrightarrow{OA}_{l_1} + \ldots + \lambda_{l_s}\overrightarrow{OA}_{l_s}.$$

Definamos mediante la función F la funcional

$$f(X) = \sum \lambda_{l_k}(F(A_{l_k}) - F(O)) + F(O).$$

Se comprueba fácilmente que la funcional f es lineal sobre $\mathfrak{A}$ y que $F \longrightarrow f$ es un isomorfismo del espacio vectorial de las funciones, definidas sobre $\mathfrak{B}$, sobre el espacio vectorial $\mathfrak{A}_f$. Por lo tanto, la dimensión de $\mathfrak{A}_f$ es igual a la potencia de la base $\mathfrak{B}$, es decir, es mayor en 1 que la dimensión de $\mathfrak{A}$.

Complementos y ejemplos

1. Sea $\mathfrak{A}$ un espacio afín sobre un espacio vectorial $\mathfrak{L} = L(\mathfrak{A})$. Para todo plano $\mathfrak{M} \subseteq \mathfrak{A}$, indiquemos por $\mathfrak{M}^\perp$ el conjunto de las funcionales lineales sobre $\mathfrak{A}$ que son iguales a 0 sobre $\mathfrak{M}$. Recíprocamente, para todo conjunto $\mathfrak{U}$ de funcionales lineales, indiquemos por $\mathfrak{U}^\top$ la variedad radical de $\mathfrak{U}$. Entonces, para un espacio $\mathfrak{A}$ de dimensión arbitraria es válida la siguiente *ley de dualidad:* para todo plano no vacío $\mathfrak{M} \subseteq \mathfrak{A}$ se tiene $(\mathfrak{M}^\perp)^\top = \mathfrak{M}$ y la dimensión de $\mathfrak{M}$ es igual a la codimensión de $\mathfrak{M}^\perp$ en el espacio lineal $\mathfrak{A}_f$ de todas las funcionales lineales sobre $\mathfrak{A}$. Recíprocamente, si un subespacio lineal $\mathfrak{U} \subset \mathfrak{A}_f$ no contiene funcionales constantes diferentes de cero, se tiene $\mathfrak{U}^\top \neq \varnothing$, $(\mathfrak{U}^\top)^\perp = \mathfrak{U}$ y la dimensión de $\mathfrak{U}$ es igual a la codimensión de $\mathfrak{U}^\top$ en $\mathfrak{A}$.

2. HOMOMORFISMOS DE ESPACIOS AFINES. Los espacios afines sobre un cuerpo conmutativo K son unas álgebras de signatura fija y, por ello, se puede hablar de *homomorfismos, congruencias* y *espacios cocientes* de los espacios afines sobre K. Es fácil ver que siendo θ una congruencia sobre un espacio afín $\mathfrak{A}$, las clases de equivalencia según θ son los planos que se obtienen uno de otro mediante traslaciones y que son, por ende, paralelos. Recíprocamente, si $\mathfrak{M}$ es un plano, todos los planos paralelos a éste, que se obtienen de $\mathfrak{M}$ mediante traslaciones, forman el sistema de clases de equivalencia según una congruencia $\theta_{\mathfrak{M}}$ sobre $\mathfrak{A}$. La dimensión del espacio cociente $\mathfrak{A}/\theta_{\mathfrak{M}}$ es igual a la codimensión de $\mathfrak{M}$.

3. De los axiomas A_1^*, A_2^* y A_3^* se ve que la clase de todos los espacios afines sobre un cuerpo conmutativo fijo K se define mediante identidades, es decir, que esta clase es *una variedad de álgebras*. Por ello se puede hablar de espacios afines *libres* sobre un cuerpo conmutativo K provistos de un sistema dado de generadores libres. Puesto que una base de un espacio de dimensión finita es un sistema de generadores del mismo, el número de sus elementos disminuido en 1 coincide con la dimensión de $\mathfrak{A}$ y como todos los espacios de una dimensión dada son isomorfos, todo espacio afín $\mathfrak{A}$ sobre K es un espacio libre y los elementos de cualquier base de $\mathfrak{A}$ son los generadores libres de $\mathfrak{A}$.

4. Indiquemos por $\mathfrak{A} \times \mathfrak{B}$ el producto cartesiano de los espacios afines $\mathfrak{A}$ y $\mathfrak{B}$ sobre un cuerpo conmutativo fijo K (comprendido como el producto cartesiano de las *álgebras* $\mathfrak{A}$ y $\mathfrak{B}$). Demuéstrese que el espacio vectorial de las traslaciones $L(\mathfrak{A} \times \mathfrak{L})$ es isomorfo a la suma directa de los espacios $L(\mathfrak{A})$ y $L(\mathfrak{B})$.

En particular, un espacio afín de dimensión finita n es isomorfo al producto cartesiano de n espacios afines de una dimensión.

5. Indiquemos por G el grupo de los automorfismos de un espacio afín $\mathfrak{A}$. Para todo punto $O \in \mathfrak{A}$ indiquemos por G_O el conjunto de todos los automorfismos de $\mathfrak{A}$ que dejan inmóvil el punto O. Sea, finalmente, D el conjunto de todas las traslaciones de $\mathfrak{A}$. Es fácil comprobar que G_O es un subgrupo de G y que D es un subgrupo abeliano invariante de G. De acuerdo con el teorema 4 del p. 29.1, tenemos la descomposición semidirecta

$$G = G_O \cdot D \quad (G_O \cap D = 1),$$

de la cual se deduce, en particular, que el grupo cociente G/D es isomorfo a G_O. Si el espacio $\mathfrak{A}$ es de una dimensión, el grupo G_O es isomorfo al grupo multiplicativo del cuerpo conmutativo K y es por lo tanto abeliano, mientras que el grupo G es metaabeliano.

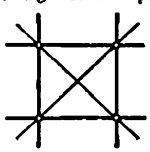


Fig. 8

6. Compruébese, en las notaciones del complemento anterior, la igualdad

$$\overrightarrow{AO} \cdot G_O \cdot \overrightarrow{OA} = \overrightarrow{OA}^{-1} \cdot G_O \cdot \overrightarrow{OA} = G_A.$$

7. Como hemos visto en el teorema 6 del p. 29.2, si la característica del cuerpo conmutativo K es diferente de 2, todo conjunto $\mathfrak{M}$ de puntos de un espacio afín sobre K que con dos cualesquiera puntos distintos suyos contiene también toda la recta que pasa por éstos, es un plano. Para cuerpos de característica 2 esto, en general, no es válido. Sea K el cuerpo conmutativo formado por dos elementos 0 y 1, sea $\mathfrak{L}$ el espacio vectorial de pares (x, y) $(x, y \in K)$ y sea $\mathfrak{A}$ el espacio afín sobre $\mathfrak{L}$ (fig. 8). Se comprueba fácilmente que $\mathfrak{A}$ tiene en total 4 puntos y 6 rectas. El conjunto $\mathfrak{M} = \{(0, 0), (0, 1), (1, 0)\}$ no es un plano, aunque con cualquier par de sus puntos contiene la recta que pasa por los mismos (que coincide con este par).

§ 30. Coordenadas afines

30.1. Coordenadas de un punto. Sea $\mathfrak{A}$ un espacio afín de dimensión finita n sobre un cuerpo conmutativo K. Según el p. 29.2, toda sucesión $(A_0, A_1, \ldots, A_n)$ formada por $n+1$ puntos linealmente independientes del espacio $\mathfrak{A}$ se llama *base* de $\mathfrak{A}$. Una sucesión $(A_0, v_1, \ldots, v_n)$, formada por un punto cualquiera $A_0 \in \mathfrak{A}$ y por unos vectores linealmente independientes $v_1, \ldots, v_n$ del espacio vectorial $L(\mathfrak{A})$, se llama *base de referencia* (hablando con más precisión, base de referencia n-dimensional) de $\mathfrak{A}$. El punto A_0 se llama origen de la base de referencia. Está claro que a toda base $(A_0, A_1, \ldots, A_n)$ le corresponde la base de referencia $(A_0, \overrightarrow{A_0A_1}, \ldots, \overrightarrow{A_0A_n})$ y que a toda base de referencia $(A_0, v_1, \ldots, v_n)$ le corresponde la base $(A_0, A_0v_1, \ldots, A_0v_n)$. Por esta razón no se hace frecuentemente diferencia entre los conceptos de base y de base de referencia.

Fijemos en el espacio $\mathfrak{A}$ una base de referencia $R = (O, \overrightarrow{OA_1}, \ldots, \overrightarrow{OA_n})$ que será llamada *base de referencia coordenada*. Para todo $X \in \mathfrak{A}$, su radio vector $\overrightarrow{OX}$ puede ser representado entonces unívocamente en

la forma

$$\overrightarrow{OX} = \xi^1 \cdot \overrightarrow{OA_1} + \ldots + \xi^n \cdot \overrightarrow{OA_n} \quad (\xi^i \in K). \tag{1}$$

Los números $\xi^1, \ldots, \xi^n$ se llaman *coordenadas del punto* X en la base de referencia R (o en la base $(O, A_1, \ldots, A_n)$) y la fila $\xi = (\xi^1, \ldots, \xi^n)$ se llama *fila coordenada* del punto X en la base de referencia señalada. Estas coordenadas se llaman a veces coordenadas afines. Al cambiar la base de referencia coordenada (incluso, por ejemplo, al cambiar el orden en el que siguen los vectores), cambia también la fila coordenada del punto. La ley de variación será considerada más tarde, en el p. 30.4.

Sean X_0, X_1 y X_2 unos puntos arbitrarios del espacio. Se dice que el punto X_1 divide al par de puntos (X_0, X_2) según la razón λ $(\lambda \in K)$, si $\overrightarrow{X_0X_1} = \lambda \overrightarrow{X_1X_2}$. De aquí se deduce, en particular, que si el punto X_1 divide al par de puntos (X_0, X_2) según una razón determinada, los tres puntos se hallan sobre una misma recta. Recíprocamente, si los puntos X_0, X_1, X_2 se hallan sobre una misma recta y $X_1 \neq X_2$, el punto X_1 divide indudablemente al par (X_0, X_2) según la razón $\lambda = \overrightarrow{X_0X_1} : \overrightarrow{X_1X_2}$, ya que los vectores $\overrightarrow{X_0X_1}$ y $\overrightarrow{X_1X_2}$ deben ser linealmente dependientes.

¿Cómo determinar —dados el número $\lambda \in K$ y las coordenadas $(\xi_0^1, \ldots, \xi_0^n)$ y $(\xi_2^1, \ldots, \xi_2^n)$ de los puntos X_0 y X_2— las coordenadas $(\xi_1^1, \ldots, \xi_1^n)$ del punto X_1 que divide al par (X_0, X_2) según la razón λ? Tenemos, por hipótesis,

$$\overrightarrow{OX_i} = \xi_i^1 \cdot \overrightarrow{OA_1} + \ldots + \xi_i^n \cdot \overrightarrow{OA_n} \quad (i = 0, 1, 2),$$

de donde

$$\overrightarrow{X_iX_{i+1}} = \overrightarrow{OX_{i+1}} - \overrightarrow{OX_i} = (\xi_{i+1}^1 - \xi_i^1) \cdot \overrightarrow{OA_1} + \ldots + (\xi_{i+1}^n - \xi_i^n) \cdot \overrightarrow{OA_n}.$$

De la relación $\overrightarrow{X_0X_1} = \lambda \overrightarrow{X_1X_2}$ resulta

$$\xi_1^j - \xi_0^j = \lambda (\xi_2^j - \xi_1^j)$$

y por ello

$$\xi_1^j = \frac{\xi_0^j + \lambda \xi_2^j}{1 + \lambda} \quad (j = 1, \ldots, n).$$

Está claro que el número dado λ debe ser diferente de -1, ya que en el caso contrario de $\overrightarrow{X_0X_1} = -\overrightarrow{X_1X_2}$ se tiene $X_0 = X_2$. Tomando $\lambda = 1$, obtenemos las fórmulas

$$\xi_1^j = \frac{\xi_0^j + \xi_2^j}{2} \quad (j = 1, \ldots, n)$$

para las coordenadas del *punto medio* X_1 del par (X_0, X_2).

Veamos ahora cómo se refleja la independencia lineal de unos puntos arbitrarios $X_1, \ldots, X_s$ del espacio $\mathfrak{A}$ en las filas coordenadas de los mismos. Sea

$$\overrightarrow{OX_i} = \xi_i^1 \cdot \overrightarrow{OA_1} + \ldots + \xi_i^n \cdot \overrightarrow{OA_n} \quad (i = 1, \ldots, s), \tag{2}$$

de modo que $(\xi_i^1, \ldots, \xi_i^n)$ es la fila coordenada del punto X_i en la

base de referencia escogida. La independencia lineal de estos puntos equivale, por definición, a la independencia lineal de los vectores $\overrightarrow{X_1X_2}, \ldots, \overrightarrow{X_1X_s}$. Tenemos de la igualdad (2)

$$\overrightarrow{X_1X_i} = \overrightarrow{OX_i} - \overrightarrow{OX_1} = (\xi_i^1 - \xi_1^1)\cdot\overrightarrow{OA_1} + \ldots + (\xi_i^n - \xi_1^n)\cdot\overrightarrow{OA_n},$$

es decir, la fila $(\xi_i^1 - \xi_1^1, \ldots, \xi_i^n - \xi_1^n)$ es la fila coordenada del vector $\overrightarrow{X_1X_i}$ en la base $\overrightarrow{OA_1}, \ldots, \overrightarrow{OA_n}$. Por esto el número máximo de puntos linealmente independientes del sistema $X_1, \ldots, X_s$, que se obtiene agregando la unidad al número máximo de vectores linealmente independientes del sistema $\overrightarrow{X_1X_2}, \ldots, \overrightarrow{X_1X_s}$, es igual a

$$1 + \text{rango} \begin{bmatrix} \xi_2^1 - \xi_1^1 & \ldots & \xi_2^n - \xi_1^n \\ \ldots & \ldots & \ldots \\ \xi_s^1 - \xi_1^1 & \ldots & \xi_s^n - \xi_1^n \end{bmatrix} = 1 + \text{rango} \, \| \xi_i^j - \xi_1^j \|.$$

Es fácil ver que

$$1 + \text{rango} \begin{bmatrix} \xi_2^1 - \xi_1^1 & \ldots & \xi_2^n - \xi_1^n \\ \ldots & \ldots & \ldots \\ \xi_s^1 - \xi_1^1 & \ldots & \xi_s^n - \xi_1^n \end{bmatrix} = \text{rango} \begin{bmatrix} 1 & \xi_1^1 & \ldots & \xi_1^n \\ 1 & \xi_2^1 & \ldots & \xi_2^n \\ \ldots & \ldots & \ldots & \ldots \\ 1 & \xi_s^1 & \ldots & \xi_s^n \end{bmatrix}.$$

Efectivamente, restando de la segunda fila, ..., de la n-ésima fila de la matriz del segundo miembro su primera fila, no alteramos su rango y, por ello,

$$\text{rango} \begin{bmatrix} 1 & \xi_1^1 & \ldots & \xi_1^n \\ 1 & \xi_2^1 & \ldots & \xi_2^n \\ \ldots & \ldots & \ldots & \ldots \\ 1 & \xi_s^1 & \ldots & \xi_s^n \end{bmatrix} = \text{rango} \begin{bmatrix} 1 & \xi_1^1 & \ldots & \xi_1^n \\ 0 & \xi_2^1 - \xi_1^1 & \ldots & \xi_2^n - \xi_1^n \\ \ldots & \ldots & \ldots & \ldots \\ 0 & \xi_s^1 - \xi_1^1 & \ldots & \xi_s^n - \xi_1^n \end{bmatrix} =$$

$$= 1 + \text{rango} \, \| \xi_i^j - \xi_1^j \|$$

que es lo que se quería demostrar.

Sea $(\xi^1, \ldots, \xi^n)$ la fila coordenada de un punto X. Con frecuencia la fila $(1, \xi^1, \ldots, \xi^n)$ suele llamarse entonces fila coordenada *ampliada* del punto X. El resultado que hemos obtenido puede ser expresado ahora en los términos siguientes: *el número máximo de puntos linealmente independientes en el sistema $X_1, \ldots, X_s$ es igual al rango de la matriz formada por las filas coordenadas ampliadas de los puntos de este sistema.*

En particular, *para que los puntos $X_1, \ldots, X_n, X_{n+1}$ de un espacio afín de n dimensiones sean linealmente independientes es necesario y suficiente que*

$$\begin{vmatrix} 1 & \xi_1^1 & \ldots & \xi_1^n \\ 1 & \xi_2^1 & \ldots & \xi_2^n \\ \ldots & \ldots & \ldots & \ldots \\ 1 & \xi_{n+1}^1 & \ldots & \xi_{n+1}^n \end{vmatrix} \neq 0, \qquad (3)$$

donde $\xi^1, \ldots, \xi^n$ *son las coordenadas del punto* X_i $(i=1, \ldots, n+1)$.

El determinante que figura en (3) a la izquierda del signo $\neq$ se llama a veces *determinante volumen* del sistema de puntos $X_1, \ldots, X_{n+1}$. Su anulación significa que los puntos $X_1, \ldots, X_{n+1}$ se hallan sobre un mismo hiperplano del espacio $\mathfrak{A}$.

Hemos introducido en el p. 29.4 el concepto de una funcional lineal $f(X)$ en un espacio afín $\mathfrak{A}$. Veamos cómo pueden expresarse los valores de esta funcional en términos de las coordenadas del punto X. Sea $(\xi^1, \ldots, \xi^n)$ la fila coordenada de un punto X en una base de referencia $R=(O, \overline{OA}_1, \ldots, \overline{OA}_n)$. De (1) y de la relación que caracteriza la linealidad de $f(X)$ obtenemos entonces

$$f(X)=\alpha_1\xi^1+\ldots+\alpha_n\xi^n+\alpha_0, \qquad (4)$$

donde $\alpha_i=f(A_i)-f(O)$ y $\alpha_0=f(O)$ son unos números fijos que dependen solamente de la funcional f y de la base de referencia R. Viceversa, tomando unos números $\alpha_0, \alpha_1, \ldots, \alpha_n \in K$ totalmente arbitrarios y definiendo la funcional f mediante la fórmula (4), podemos comprobar fácilmente que f es lineal sobre $\mathfrak{A}$.

Se dice que *el polinomio* $\alpha_1\xi^1+\ldots+\alpha_n\xi^n+\alpha_0$ *en las variables* $\xi^1, \ldots, \xi^n$ *representa la funcional* f *en la base de referencia* R. De lo expuesto se ve que la correspondencia entre los polinomios lineales en n variables $\xi^1, \ldots, \xi^n$ y las funcionales lineales sobre $\mathfrak{A}$ es biyectiva; además, si la funcional f está representada por el polinomio $\alpha_1\xi^1+\ldots+\alpha_n\xi^n+\alpha_0$ y la funcional g por el polinomio $\beta_1\xi^1+\ldots+\beta_n\xi^n+\beta_0$, se tiene

$$\lambda f(X)+\mu g(X)=(\lambda\alpha_1+\mu\beta_1)\xi^1+\ldots+(\lambda\alpha_n+\mu\beta_n)\xi^n+\lambda\alpha_0+\mu\beta_0. \qquad (5)$$

En otras palabras, poniendo en correspondencia a toda funcional lineal sobre $\mathfrak{A}$ el polinomio en $\xi^1, \ldots, \xi^n$ que la representa, obtenemos *una aplicación isomorfa* del espacio vectorial $\mathfrak{A}_f$ de todas las funcionales lineales, definidas sobre $\mathfrak{A}$, sobre el espacio vectorial de todos los polinomios lineales (no homogéneos) en $\xi^1, \ldots \xi^n$ con coeficientes del cuerpo conmutativo principal K.

De (5) se ve que las operaciones de adición de las funcionales lineales y de multiplicación de las mismas por un número se reducen a las operaciones correspondientes con las filas de coeficientes de los polinomios lineales. De aquí deducimos, en particular, que el número máximo de funcionales lineales linealmente independientes, contenidas en el sistema $f_1(X), \ldots, f_s(X)$, es igual al rango de la matriz formada por los coeficientes de las formas lineales que representan a las funcionales señaladas.

30.2. Ecuaciones de planos. Aceptaremos que en un espacio $\mathfrak{A}$ de n dimensiones sobre un cuerpo conmutativo K se ha fijado una base de referencia $R=(O, \overline{OA}_1, \ldots, \overline{OA}_n)$ y en lo sucesivo enten-

deremos por coordenadas de los puntos de $\mathfrak{A}$ las coordenadas de éstos en la base de referencia R.

Sea $P(\xi^1, \ldots, \xi^n)$ *una condición* que relaciona unos números arbitrarios $\xi^1, \ldots, \xi^n$ de K. Luego, para todo sistema concreto de valores de las variables $\xi^1, \ldots, \xi^n \in K$ la condición P puede ser verídica o falsa. Se dice que *la condición P define en el espacio $\mathfrak{A}$ un conjunto de puntos $\mathfrak{M}$, si cualquier punto $X \in \mathfrak{A}$ pertenece a $\mathfrak{M}$ cuando, y sólo cuando, las coordenadas $(\xi^1, \ldots, \xi^n)$ del punto X satisfacen la condición P.*

De aquí resulta, en particular, que si las condiciones $P(\xi^1, \ldots, \xi^n)$ y $Q(\xi^1, \ldots, \xi^n)$ definen en el espacio $\mathfrak{A}$ los conjuntos respectivos $\mathfrak{M}$ y $\mathfrak{N}$, la conjunción de estas condiciones $P(\xi^1, \ldots, \xi^n)$ *y* $Q(\xi^1, \ldots, \xi^n)$ define la intersección de los conjuntos señalados, la disjunción P *o* Q define la unión $\mathfrak{M} \cup \mathfrak{N}$ de los mismos y la negación *no* P define el complemento $C\mathfrak{M} = \mathfrak{A} \setminus \mathfrak{M}$.

Sea $f(\xi^1, \ldots, \xi^n)$ una función que está definida sobre K y toma valores en K. La condición de tipo

$$f(\xi^1, \ldots, \xi^n) = 0 \tag{1}$$

se llama *ecuación* en las variables $\xi^1, \ldots, \xi^n$ y la conjunción de las ecuaciones

$$f_i(\xi^1, \ldots, \xi^n) = 0 \quad (i = 1, 2, \ldots, s) \tag{2}$$

se llama *sistema de ecuaciones* en las variables $\xi^1, \ldots, \xi^n$. Si el sistema de ecuaciones (2) define en el espacio $\mathfrak{A}$ un conjunto $\mathfrak{M}$, se dice también que (2) es el sistema de ecuaciones para el conjunto $\mathfrak{M}$. El conjunto $\mathfrak{M}$ de puntos definido por el sistema de ecuaciones (2) es, según la observación hecha anteriormente, la intersección de los conjuntos definidos por cada una de las ecuaciones del sistema (2) por separado.

En un cuerpo conmutativo arbitrario la condición $\alpha\beta = 0$ equivale a la disjunción $\alpha = 0$ *o* $\beta = 0$. Por esto, si las ecuaciones

$$f(\xi^1, \ldots, \xi^n) = 0 \quad \text{y} \quad g(\xi^1, \ldots, \xi^n) = 0$$

definen (por separado) en el espacio $\mathfrak{A}$ los conjuntos respectivos $\mathfrak{M}$ y $\mathfrak{N}$, la ecuación

$$f(\xi^1, \ldots, \xi^n) \cdot g(\xi^1, \ldots, \xi^n) = 0 \tag{3}$$

define en $\mathfrak{A}$ la unión de los conjuntos $\mathfrak{M}$ y $\mathfrak{N}$.

Un conjunto $\mathfrak{M}$ de puntos de un espacio afín $\mathfrak{A}$ se llama *hipersuperficie algebraica*, si existe un polinomio $f(\xi^1, \ldots, \xi^n)$ en las variables $\xi^1, \ldots, \xi^n$ con coeficientes del cuerpo conmutativo K tal que $\mathfrak{M}$ se define por la ecuación (1).

El grado del polinomio f respecto al conjunto de las variables $\xi^1, \ldots, \xi^n$ se llama *grado* de la hipersuperficie $\mathfrak{M}$. Puesto que un mismo conjunto $\mathfrak{M}$ puede tener varias ecuaciones diferentes, una misma hipersuperficie algebraica puede tener diferentes grados. Es

fácil darse cuenta, en particular, de que si una hipersuperficie tiene el grado n, cualquier número mayor también es grado de ella. El menor de los grados de una hipersuperficie algebraica dada se llama *orden* de la misma.

Como el grado de un producto de polinomios es igual a la suma de los grados de los factores, de la fórmula (3) se deduce que la unión de unas hipersuperficies algebraicas de ordenes s y t es de nuevo una superficie algebraica de orden no mayor que $s+t$.

Se dice que una hipersuperficie algebraica $\mathfrak{M}$ se descompone, si es la unión de dos hipersuperficies algebraicas no vacías diferentes de la dada.

La intersección de un número finito de hipersuperficies algebraicas se llama *variedad algebraica*.

Los conceptos de hipersuperficie algebraica y de su orden han sido definidos mediante el tipo de las ecuaciones a las que satisfacen las coordenadas de un punto arbitrario de la hipersuperficie. Al cambiar la base de referencia coordenada, también cambiarán las coordenadas del punto y con ellas las ecuaciones del conjunto considerado. En lugar de hipersuperficie algebraica de orden s, sería más correcto hablar por esto *de una hipersuperficie algebraica de orden s en una base de referencia coordenada dada* y, análogamente, *de la descomposición en la base de referencia coordenada dada*, etc. Sin embargo, demostraremos en adelante que la propiedad de un conjunto de ser una hipersuperficie algebraica de un orden dado no depende, de hecho, de la selección de la base de referencia coordenada.

Las propiedades de las variedades algebraicas arbitrarias constituyen el tema de estudio de una asignatura especial, la Geometría algebraica. En este libro serán estudiadas solamente las hipersuperficies de primer orden (los hiperplanos) y sus intersecciones (los planos).

Es poco cómodo que las definiciones de *algebraicidad* y de *orden* dependan, como hemos señalado, de la selección de la base de referencia coordenada. Para deshacerse de esta dependencia se emplea el siguiente procedimiento. Sea $f(X)$ una funcional dada en un espacio afín $\mathfrak{A}$ de dimensión n. Escojamos en $\mathfrak{A}$ una base de referencia coordenada cualquiera $R=(O, \overrightarrow{OA}_1, \ldots, \overrightarrow{OA}_n)$ y pongamos en correspondencia a toda sucesión $\xi^1, \ldots, \xi^n$ de números de K el número $f(\xi^1, \ldots, \xi^n)$ igual a $f(X)$, donde X es el punto de coordenadas $\xi^1, \ldots, \xi^n$. Resulta así que a toda funcional se le pone en correspondencia una función en n variables $\xi^1, \ldots, \xi^n$ que *representa* a la funcional f en la base de referencia coordenada dada R. Recíprocamente, definiendo para la función $f(\xi^1, \ldots, \xi^n)$ la funcional $f(X)$ mediante la igualdad

$$f(X)=f(\xi^1, \ldots, \xi^n),$$

vemos que cualquier función en n variables representa a una funcional en la base de referencia señalada. Está claro que *la variedad radical de la funcional* $f(X)$, compuesta de aquellos puntos $X \in \mathfrak{A}$ para los cuales $f(X) = 0$, coincide con el conjunto que la ecuación $f(\xi^1, \ldots, \xi^n) = 0$ define en la base de referencia coordenada dada. Sin embargo, en la definición del concepto de una funcional y de su variedad radical no figuran las bases de referencia coordenadas y, por ello, es preferible representar las ecuaciones de un conjunto de puntos en la forma $f(X) = 0$, donde f es una funcional. ¿Cómo entonces caracterizar aquellas funcionales cuyos valores se representan por polinomios? La respuesta es la siguiente. Hemos visto ya en el p. 30.1 que los polinomios *lineales* representan a las funcionales lineales. Consideremos ahora las funcionales $F(X_1, \ldots, X_s)$ en s puntos variables. Una funcional $F(X_1, \ldots, X_s)$ se llama *lineal respecto al i-ésimo argumento* X_i, si se convierte en una funcional lineal en X_i para cualesquiera valores fijos de las restantes variables. Una funcional $F(X_1 \ldots, X_s)$ se llama *polilineal* si es lineal respecto a cada uno de sus argumentos. Hemos llegado ahora al momento central: una funcional $f(X)$ en una variable X se llama *funcional* (o *forma*) *de orden* s (o *de grado* s), si existe una funcional polilineal $F(X_1, \ldots, X_s)$ en s variables tal que

$$f(X) \equiv F(X, \ldots, X).$$

Se comprueba fácilmente que habiendo sido fijada una base de referencia coordenada cualquiera, la funcional $f(X)$ es de orden s cuando, y sólo cuando, sus valores se representan por un polinomio adecuado de orden s en las coordenadas del punto X. La demostración la omitimos aquí debido a su evidencia.

Después de estas consideraciones generales pasamos ahora al problema principal de este parágrafo que es el estudio de las ecuaciones de los planos.

Supongamos, pues, que en un espacio afín dado de dimensión finita n se ha fijado una base de referencia coordenada arbitraria $R = (O, \overline{OA}_1, \ldots, \overline{OA}_n)$. Una ecuación de primer grado en las variables $\xi^1, \ldots, \xi^n$ es una ecuación de tipo

$$\alpha_1\xi^1 + \ldots + \alpha_n\xi^n = \beta, \tag{4}$$

donde al menos uno de los coeficientes $\alpha_1, \ldots, \alpha_n$ es diferente de cero. Introduciendo la funcional

$$f(X) = \alpha_1\xi^1 + \ldots + \alpha_n\xi^n - \beta,$$

vemos que el conjunto de puntos $\mathfrak{M}$ representado por la ecuación (4) es la variedad radical de la funcional $f(X)$. La funcional $f(X)$ es, según el p. 30.1, lineal y no es constante, de modo que $\mathfrak{M}$ es un hiperplano en $\mathfrak{A}$. Luego, *una ecuación lineal arbitraria en unas variables* $\xi^1, \ldots, \xi^n$ *representa un hiperplano en un espacio afín.*

Consideremos un sistema cualquiera de ecuaciones lineales

$$\left.\begin{array}{l}\alpha_1^1\xi^1+\alpha_2^1\xi^2+\ \ldots+\alpha_n^1\xi^n=\beta^1,\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ \alpha_1^s\xi^1+\alpha_2^s\xi^2+\ldots+\alpha_n^s\xi^n=\beta^s.\end{array}\right\} \tag{5}$$

Indicando por $\mathfrak{M}_i$ el conjunto de los puntos cuyas coordenadas satisfacen la i-ésima ecuación de (5), vemos que el sistema de ecuaciones (5) define la intersección de hiperplanos $\mathfrak{M}=\mathfrak{M}_1\cap\ldots\cap\mathfrak{M}_s$. Según el p. 29.2, esta intersección o bien es vacía o bien es un plano cuya codimensión no pasa de s, de modo que la dimensión de $\mathfrak{M}$ es no menor que $n-s$. ¿Cuál es el valor exacto de la dimensión de $\mathfrak{M}$?

Introduzcamos las funcionales lineales

$$f_i(X)=\alpha_1^i\xi^1+\ldots+\alpha_n^i\xi^n-\beta^i \quad (i=1,\ \ldots,\ s).$$

El plano $\mathfrak{M}$ es la variedad radical del sistema de estas funcionales. Si $\mathfrak{M}\neq\varnothing$, la dimensión de $\mathfrak{M}$ es igual, por lo visto en el teorema 2 del p. 29.4, al número máximo de funcionales linealmente independientes en el sistema $f_1,\ \ldots,\ f_s$, es decir (véase el p. 29.4), la dimensión de $\mathfrak{M}$ es igual al rango de la matriz $\|\alpha_j^i\|$ formada por los coeficientes de las variables del sistema de ecuaciones (5).

Resta aclarar en qué caso $\mathfrak{M}=\varnothing$ y en qué caso $\mathfrak{M}\neq\varnothing$, es decir, bajo qué condiciones el sistema (5) es compatible. Pero la respuesta a esta última pregunta viene dada por el teorema de Kronecker—Capelli (véase el p. 5.3): para que el sistema de ecuaciones (5) sea compatible es necesario y suficiente que el rango de la matriz principal de este sistema coincida con el rango de su matriz ampliada. Hemos obtenido de esta forma el teorema siguiente:

TEOREMA. *En un espacio afín $\mathfrak{A}$ de n dimensiones todo plano $(n-r)$-dimensional $(0\leqslant r\leqslant n)$ $\mathfrak{M}$ puede ser representado, en cualquier base de referencia coordenada, por un sistema de ecuaciones de tipo*

$$\alpha_1^i\xi^1+\alpha_2^i\xi^2+\ldots+\alpha_n^i\xi^n=\beta^i \quad (i=1,\ \ldots,\ r) \tag{6}$$

tal que el rango de la matriz principal $\|\alpha_j^i\|$ es igual a r. Un sistema arbitrario de ecuaciones lineales de tipo (4) *representa en $\mathfrak{A}$ un plano $(n-r)$-dimensional siempre que los rangos de las matrices principal y ampliada del sistema* (4) *coincidan y sean iguales a r. En cambio, si dichos rangos son diferentes, el sistema* (4) *es incompatible y, por consiguiente, representa el plano vacío.*

Consideremos el problema siguiente. Dadas las filas coordenadas $(\beta_i^1,\ \ldots,\ \beta_i^n)$ de unos puntos B_i $(i=0,\ 1,\ \ldots,\ s)$ de un espacio $\mathfrak{A}$, hallar la ecuación del plano mínimo $\mathfrak{M}=B_0\vee B_1\vee\ldots\vee B_s$ que pasa por los puntos dados.

Resolvámoslo primero en la forma vectorial. El plano $\mathfrak{M}$ se compone, según el p. 29.2, de los puntos $X \in \mathfrak{A}$ tales que el vector $\overline{B_0X}$ puede ser representado en la forma

$$\overline{B_0X} = \lambda_1 \overline{B_0B_1} + \ldots + \lambda_s \overline{B_0B_s} \quad (\lambda_1, \ldots, \lambda_s \in K),$$

es decir,

$$\overline{OX} = \lambda_1 (\overline{OB_1} - \overline{OB_0}) + \ldots + \lambda_s (\overline{OB_s} - \overline{OB_0}) + \overline{OB_0}. \qquad (7)$$

La ecuación (7) se llama a veces ecuación de un plano *en la forma vectorial paramétrica*. Introduciendo aquí en lugar de los radios vectores $\overline{OX}$ y $\overline{OB_j}$ sus filas coordenadas correspondientes, obtenemos las relaciones

$$\xi^i = (\beta_1^i - \beta_0^i)\lambda_1 + \ldots + (\beta_s^i - \beta_0^i)\lambda_s + \beta_0^i \quad (i = 1, \ldots, n). \qquad (8)$$

Poniendo $\beta_j^i - \beta_0^i = \gamma_j^i$, podemos representar estas relaciones en la forma

$$\xi^i = \gamma_1^i\lambda_1 + \ldots + \gamma_s^i\lambda_s + \beta_0^i \quad (i = 1, \ldots, n), \qquad (9)$$

que expresa las coordenadas de un punto arbitrario $X \in \mathfrak{M}$ en términos de los parámetros independientes auxiliares $\lambda_1, \ldots, \lambda_s$. Las ecuaciones (9) se llaman *ecuaciones coordenadas paramétricas* del plano $\mathfrak{M}$. Puesto que $\mathfrak{M} = B_0 \vee B_1 \vee \ldots \vee B_s$, la dimensión de $\mathfrak{M}$ es igual al número máximo de puntos linealmente independientes que figuran en el sistema $B_0, B_1, \ldots, B_s$ disminuido en 1, es decir, es igual al rango de la matriz $C = \| \gamma_j^i \|$ formada por los coeficientes de las variables $\lambda_1, \ldots, \lambda_s$ de las ecuaciones paramétricas.

Para obtener de las ecuaciones paramétricas (9) de un plano $\mathfrak{M}$ sus ecuaciones generales de tipo (5), es suficiente eliminar del sistema (9) los parámetros $\lambda_1, \ldots, \lambda_s$. Esto se puede hacer, por ejemplo, del modo siguiente. Buscamos el rango de la matriz $\| \gamma_j^i \|$. Supongamos que éste es igual a r; luego, entre las ecuaciones (9) existen unas ecuaciones i_1-ésima, i_2-ésima, ..., i_r-ésima y entre los parámetros $\lambda_1, \ldots \lambda_s$ existen unos parámetros $\lambda_{j_1}, \ldots, \lambda_{j_r}$ $(i_1 < \ldots < i_r;\ j_1 < \ldots < j_r)$ tales que $\det \left\| \gamma_{j_\beta}^{i_\alpha} \right\| \neq 0$. Resolviendo estas ecuaciones respecto a las incógnitas $\lambda_{j_1}, \ldots, \lambda_{j_r}$, obtendremos para éstas unas expresiones lineales en términos de $\xi_1, \ldots, \xi_n$ y de los restantes parámetros λ_j $(j \notin \{j_1, \ldots, j_r\})$. Introduzcamos ahora los valores obtenidos de los parámetros en cada una de las ecuaciones restantes del sistema (9). Obtendremos así unas relaciones lineales entre las variables $\xi_1, \ldots, \xi_n$ y los parámetros λ_j, $j \notin \{j_1, \ldots, j_r\}$. Los parámetros λ_j aparecerán, de hecho, en estas relaciones con coeficientes nulos (debido a que r es el rango de la matriz $\| \gamma_j^i \|$) y, por ello, las relaciones obtenidas serán unas relaciones de tipo (5) que representarán el plano $\mathfrak{M}$.

En los razonamientos realizados hemos supuesto que $0 \leqslant r < n$. Si $r = n$, el plano $\mathfrak{M}$ pasará por $n+1$ puntos linealmente independientes y coincidirá con todo el espacio $\mathfrak{A}$; si se quiere, su «ecuación general» puede ser representada en la forma

$$0 \cdot \xi^1 + \ldots + 0 \cdot \xi^n = 0. \tag{10}$$

Consideremos el problema recíproco: ¿cómo hallar las ecuaciones paramétricas de un plano $\mathfrak{M}$ dado por unas ecuaciones generales de tipo (5)? Supongamos que la dimensión de $\mathfrak{M}$ es $r < n$. El sistema (5) contiene entonces sólo r ecuaciones independientes y los coeficientes que tienen en estas ecuaciones unas r variables forman una matriz de determinante diferente de 0. Supongamos, para concretar, que son independientes las r primeras ecuaciones y que es diferente de cero el determinante formado por los coeficientes de las variables $\xi^1, \ldots, \xi^r$. Resolviendo las r primeras ecuaciones respecto a las variables $\xi^1, \ldots, \xi^r$, obtendremos entonces un sistema de ecuaciones de tipo

$$\xi^i = \gamma^i_{r+1}\xi^{r+1} + \ldots + \gamma^i_n \xi^n + \beta^i \quad (i = 1, \ldots, r) \tag{11}$$

equivalente al sistema (9). Está claro que el sistema (11) es equivalente a las ecuaciones paramétricas

$$\begin{aligned} &\xi^i = \gamma^i_{r+1}\lambda_1 + \ldots + \gamma^i_n \lambda_{n-r} + \beta^i \quad (i = 1, \ldots, r), \\ &\xi^j = \lambda_{j-r} \qquad\qquad\qquad\qquad\quad (j = r+1, \ldots, n). \end{aligned}$$

Hemos considerado detalladamente el problema de la determinación de las ecuaciones del plano mínimo que pasa por los puntos dados. Está claro que dicho problema representa un caso particular de un problema más general sobre la determinación del plano mínimo que pasa por los planos dados $\mathfrak{M}$ y $\mathfrak{N}$. Sin embargo, este problema más general puede ser reducido de un modo formal al primer problema. En efecto, los planos $\mathfrak{M}$ y $\mathfrak{N}$ pueden ser representados en la forma

$$\begin{aligned} &\mathfrak{M} = B_0 \vee B_1 \vee \ldots \vee B_s \quad (\dim \mathfrak{M} = s), \\ &\mathfrak{N} = C_0 \vee C_1 \vee \ldots \vee C_t \quad (\dim \mathfrak{N} = t), \end{aligned}$$

donde $B_0, B_1, \ldots, B_s$ son unos puntos linealmente independientes de $\mathfrak{M}$ y $C_0, C_1, \ldots, C_t$ son unos puntos linealmente independientes de $\mathfrak{N}$. Entonces

$$\mathfrak{M} \vee \mathfrak{N} = B_0 \vee \ldots \vee B_s \vee C_0 \vee \ldots \vee C_t$$

y el problema ha quedado reducido a la determinación del plano mínimo que pasa por los puntos $B_0, \ldots, B_s, C_0, \ldots, C_t$. ¿Cómo determinar los puntos $B_0, \ldots, B_s$ si el plano $\mathfrak{M}$ viene dado por sus ecuaciones generales de tipo (5)? Uno de los métodos (que no es el más breve) es el siguiente. Reducimos el sistema (5) a la

forma (11). Indudablemente podemos tomar entonces para los puntos B_i los puntos de las filas coordenadas siguientes

$$\begin{aligned}
(B_0) &= (\beta^1, \quad \beta^2, \quad \ldots, \beta^r, \quad 0, 0, \ldots, 0),\\
(B_1) &= (\beta^1 + \gamma^1_{r+1}, \beta^2 + \gamma^2_{r+1}, \ldots, \beta^r + \gamma^r_{r+1}, 1, 0, \ldots, 0),\\
(B_2) &= (\beta^1 + \gamma^1_{r+2}, \beta^2 + \gamma^2_{r+2}, \ldots, \beta^r + \gamma^r_{r+2}, 0, 1, \ldots, 0),\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
(B_s) &= (\beta^1 + \gamma^1_n, \beta^2 + \gamma^2_n, \ldots, \beta^r + \gamma^r_n, 0, 0, \ldots, 1)\\
&\qquad (s + r = n).
\end{aligned}$$

Consideremos otro problema particular. Supongamos que en un espacio $\mathfrak{A}$ se ha escogido una base de referencia coordenada $R = (O, \overrightarrow{OA_1}, \ldots, \overrightarrow{OA_n})$. Los planos de tipo $O \vee A_{i_1} \vee \ldots \vee A_{i_r}$ $(1 \leqslant i_1 < \ldots < i_r \leqslant n)$ se llaman entonces *planos coordenados r-dimensionales*. Poniendo

$$\mathfrak{P}_i = O \vee A_1 \vee \ldots \vee A_{i-1} \vee A_{i+1} \vee \ldots \vee A_n, \tag{12}$$

vemos que $\mathfrak{P}_1, \mathfrak{P}_2, \ldots, \mathfrak{P}_n$ son todos *los hiperplanos coordenados* posibles. De (12) resulta que un punto X de coordenadas $(\xi^1, \ldots, \xi^n)$ pertenece a $\mathfrak{P}_i$ cuando, y sólo cuando,

$$\xi^i = 0. \tag{13}$$

En otras palabras, *la ecuación* (13) *es la ecuación del hiperplano* $\mathfrak{P}_i$. El plano coordenado $O \vee A_{i_1} \vee \ldots \vee A_{i_r} = \mathfrak{M}$ es la intersección de todos los hiperplanos $\mathfrak{P}_j$ que tienen el índice diferente de $i_1, \ldots, i_r$. Por esto las ecuaciones del plano $\mathfrak{M}$ pueden ser representadas en la forma

$$\xi^\alpha = \xi^\beta = \ldots = \xi^\gamma = 0,$$

donde $\alpha, \beta, \ldots, \gamma$ es la sucesión de aquellos números naturales del conjunto $1, 2, \ldots, n$ que no figuran en el conjunto $\{i_1, \ldots, i_r\}$. Por ejemplo, las ecuaciones del i-ésimo eje coordenado $O \vee A_i$ pueden ser representadas en la forma

$$\xi^1 = \ldots = \xi^{i-1} = \xi^{i+1} = \ldots = \xi^n = 0.$$

A título de ilustración consideremos el siguiente ejemplo numérico. En el espacio afín R^3 corriente real de tres dimensiones está dada mediante dos ecuaciones

$$\left.\begin{aligned} \xi^1 + 2\xi^2 - \xi^3 &= 1,\\ \xi^1 + 2\xi^2 - 2\xi^3 &= 2 \end{aligned}\right\} \tag{14}$$

una variedad lineal $\mathfrak{M}$.

¿Cuál es la dimensión de $\mathfrak{M}$? Componiendo la matriz ampliada de este sistema

$$\begin{bmatrix} 1 & 2 & -1 & 1 \\ 1 & 2 & -2 & 2 \end{bmatrix},$$

vemos que contiene un determinante de segundo orden diferente de cero, cuyos elementos están indicados con las cifras gruesas. Es más, este determinante pertenece a la matriz principal del sistema. Por esto el rango de la matriz ampliada, que es igual a dos, coincide con el rango de la matriz principal. De aquí sacamos la conclusión de que la codimensión de $\mathfrak{M}$ es igual a 2 y de que la dimensión de $\mathfrak{M}$ es igual a 1, es decir, $\mathfrak{M}$ es una recta.

Resolviendo el sistema (14) respecto a ξ^1 y ξ^3, encontramos

$$\begin{aligned} \xi^3 &= -1, \\ \xi^1 &= -2\xi^2. \end{aligned} \tag{15}$$

Podemos aceptar aquí que ξ^2 es un parámetro, de modo que (15) puede ser considerado como las ecuaciones paramétricas de la recta $\mathfrak{M}$. Dando al parámetro ξ^2 los valores 0 y 1, obtenemos de (15) que los puntos de coordenadas (0, 0, —1) y (— 2, 1, — 1) se hallan sobre la recta $\mathfrak{M}$.

Consideremos otro ejemplo. ¿Qué conjuntos de puntos del espacio R^3 son representados por las ecuaciones

a) $\xi^1\xi^2\xi^3 = 0$,

b) $(\xi^1)^2 - \xi^1\xi^2 + \xi^1\xi^3 + \xi^1 - \xi^2 + \xi^3 = 0$?

Está claro que en el caso a) el conjunto incógnito es la colección de los tres hiperplanos coordenados de ecuaciones 1) $\xi^1 = 0$, 2) $\xi^2 = 0$ y 3) $\xi^3 = 0$. En el caso b), el primer miembro se descompone en los factores $\xi^1 + 1$ y $\xi^1 - \xi^2 + \xi^3$ y, por ello, el conjunto incógnito es la colección de los hiperplanos de ecuaciones $\xi^1 + 1 = 0$ y $\xi^1 - \xi^2 + \xi^3 = 0$.

30.3. Ecuaciones de hiperplanos y de rectas. Queremos examinar aquí más detalladamente las ecuaciones de los hiperplanos y de las rectas, así como las condiciones de paralelismo de los mismos, expresadas en la forma coordenada.

TEOREMA 1. *Todo hiperplano de un espacio afín de n dimensiones puede ser representado mediante una ecuación de primer grado*

$$\alpha_1\xi^1 + \ldots + \alpha_n\xi^n + \alpha_0 = 0. \tag{1}$$

Para que un hiperplano representado por la ecuación (1) *coincida con un hiperplano representado por la ecuación*

$$\beta_1\xi^1 + \ldots + \beta_n\xi^n + \beta_0 = 0, \tag{2}$$

es necesario y suficiente que sean proporcionales los coeficientes correspondientes:

$$\frac{\alpha_1}{\beta_1} = \ldots = \frac{\alpha_n}{\beta_n} = \frac{\alpha_0}{\beta_0}. \tag{3}$$

Los hiperplanos representados por las ecuaciones (1) *y* (2) *son paralelos cuando, y sólo cuando, sus respectivos coeficientes*

principales son proporcionales:

$$\frac{\alpha_1}{\beta_1} = \ldots = \frac{\alpha_n}{\beta_n}, \tag{4}$$

El teorema contiene tres afirmaciones. La primera ha sido ya demostrada en el p. 30.2. Para que los hiperplanos (1) y (2) coincidan es necesario y suficiente que el conjunto de las ecuaciones (1) y (2) defina un hiperplano. Esto equivale, según el teorema del p. 30.2, a que

$$\text{rango} \begin{bmatrix} \alpha_1 & \ldots & \alpha_n & \alpha_0 \\ \beta_1 & \ldots & \beta_n & \beta_0 \end{bmatrix} = 1$$

y esto significa precisamente que tiene lugar (3).

Los hiperplanos (1) y (2) son paralelos, si coinciden, y entonces tenemos (3), o si no se intersecan. En el último caso las ecuaciones (1) y (2) deben ser incompatibles y por esto, según el teorema de Kronecker—Capelli, se tiene

$$\text{rango} \begin{bmatrix} \alpha_1 & \ldots & \alpha_n \\ \beta_1 & \ldots & \beta_n \end{bmatrix} = 1,$$

es decir, se tiene (4). Luego, la condición (4) es verídica en ambos casos.

Consideremos el problema: *dada la ecuación* (1) *de un hiperplano* $\mathfrak{P}$ *y las coordenadas* $\xi_0^1, \ldots, \xi_0^n$ *de un punto* A, *hallar la ecuación del hiperplano que es paralelo a* $\mathfrak{P}$ *y que pasa por el punto* A.

Sea (2) la ecuación del hiperplano $\mathfrak{P}'$ que buscamos. Puesto que este hiperplano es paralelo al hiperplano (1), deben cumplirse las condiciones (4). Por esto multiplicando la ecuación (2) por un coeficiente de proporcionalidad adecuado, podemos representar la ecuación del hiperplano que buscamos $\mathfrak{P}'$ en la forma

$$\alpha_1 \xi^1 + \ldots + \alpha_n \xi^n + \beta_0' = 0. \tag{5}$$

El punto A se halle, por hipotesis, sobre $\mathfrak{P}'$. Por ello

$$\alpha_1 \xi_0^1 + \ldots + \alpha_n \xi_0^n + \beta_0' = 0. \tag{6}$$

Restando de (5) término por término (6), obtenemos la ecuación deseada

$$\alpha_1 (\xi^1 - \xi_0^1) + \ldots + \alpha_n (\xi^n - \xi_0^n) = 0.$$

Está claro que el hiperplano definido por la ecuación (1) pasa por el origen $(0, \ldots, 0)$ de la base de referencia coordenada cuando, y sólo cuando, $\alpha_0 = 0$. Consideremos el caso en que $\alpha_0 \neq 0$. Dividiendo por $-\alpha_0$ todos los coeficientes de la ecuación (1), podemos reducirla a la forma

$$\frac{\xi^1}{\gamma_1} + \ldots + \frac{\xi^n}{\gamma_n} = 1. \tag{7}$$

Los números $\gamma_1, \ldots, \gamma_n$ admiten una interpretación geométrica simple. Efectivamente, busquemos el punto de intersección del hiperplano (7) y del eje coordenado $O \vee A_i$. Las ecuaciones de este eje son

$$\xi^1 = \xi \ldots = \xi^{i-1} = \xi^{i+1} = \ldots = \xi^n = 0. \tag{8}$$

Resolviendo conjuntamente (7) y (8), obtenemos $\xi^i = \gamma_i$. Por consiguiente, el hiperplano (7) intercepta en el eje $O \vee A_i$ el vector $\gamma_i \overrightarrow{OA_i}$. Por esta razón la ecuación (7) suele llamarse generalmente *ecuación segmentaria de un hiperplano.*

Examinemos también el siguiente problema: *¿bajo qué condiciones la ecuación* (1) *representa un hiperplano que pasa por el plano coordenado* $O \vee A_1 \vee \ldots \vee A_r$ $(1 \leqslant r < n)$? Este plano esta formado por los puntos de coordenadas $(\xi^1, \ldots, \xi^r, 0, \ldots, 0)$. Introduciendo estas coordenadas en la ecuación (1), obtenemos $\alpha_1 \xi^1 + \ldots + \alpha_r \xi^r + \alpha_0 = 0$. Esta relación debe satisfacerse para cualesquiera valores de los números $\xi^1, \ldots, \xi^r$ del cuerpo conmutativo K. Por consiguiente, $\alpha_1 = \ldots = \alpha_r = \alpha_0 = 0$. Análogamente se comprueba que el hiperplano (1) pasa por el plano coordenado r-dimensional $O \vee A_{i_1} \vee \ldots \ldots \vee A_{i_r}$, cuando, y sólo cuando, $\alpha_{i_1} = \ldots = \alpha_{i_r} = \alpha_0 = 0$.

Manteniendo invariables en la ecuación (1) los coeficientes principales $\alpha_1, \ldots, \alpha_n$ y dando diferentes valores al término independiente α_0, obtenemos ecuaciones de hiperplanos paralelos. Por esto, *si en la ecuación* (1) *se tiene* $\alpha_{i_1} = \ldots = \alpha_{i_r} = 0$, *mientras que el término independiente* α_0 *es arbitrario, el hiperplano* (1) *es paralelo al plano coordenado* $O \vee A_{i_1} \vee \ldots \vee A_{i_r}$. En particular, el hiperplano (1) es paralelo al eje coordenado $O \vee A_i$ cuando, y sólo cuando, $\alpha_i = 0$, es decir, cuando la ecuación (1) no contiene explícitamente la i-ésima coordenada.

Como sabemos, por cualesquiera n puntos linealmente independientes $B_1, \ldots, B_n$ de un espacio afín de n dimensiones pasa un hiperplano $B_1 \vee \ldots \vee B_n$ y sólo uno. ¿Cómo escribir la ecuación de este hiperplano, si se conocen las filas coordenadas de los puntos $B_1, \ldots, B_n$? Supongamos que $(\beta_1^1, \ldots, \beta_1^n), \ldots, (\beta_n^1, \ldots, \beta_n^n)$ son las filas coordenadas de los puntos $B_1, \ldots, B_n$. Para que un punto arbitrario X de coordenadas $\xi^1, \ldots, \xi^n$ se halle sobre el hiperplano $B_1 \vee \ldots \vee B_n$ es necesario y suficiente que el sistema de puntos $X, B_1, \ldots, B_n$ sea linealmente dependiente. Según el p. 30.1, el sistema de puntos señalado será linealmente dependiente cuando, y sólo cuando, las coordenadas de estos puntos satisfacen la condición

$$\begin{vmatrix} 1 & \xi^1 & \ldots & \xi^n \\ 1 & \beta_1^1 & \ldots & \beta_1^n \\ \ldots & \ldots & \ldots & \ldots \\ 1 & \beta_n^1 & \ldots & \beta_n^n \end{vmatrix} = 0. \tag{9}$$

Desarrollando el determinante según los elementos de la primera fila, vemos que la ecuación (9) es lineal respecto a las variables $\xi^1, \ldots, \xi^n$ y, por ello, representa el plano deseado.

El término independiente de la ecuación (9) es igual a $\det \|\beta_i^j\|$. Por consiguiente, el hiperplano que pasa por los puntos linealmente independientes $B_1, \ldots, B_n$ pasará por el origen de coordenadas cuando, y sólo cuando, $\det \|\beta_i^j\| = 0$.

Consideremos ahora más detalladamente las ecuaciones de las rectas. Supongamos dado en un espacio afín $\mathfrak{A}$ de n dimensiones un par de puntos diferentes $B_0(\beta_0^1, \ldots, \beta_0^n)$ y $B_1(\beta_1^1, \ldots, \beta_1^n)$, donde se indican entre paréntesis las coordenadas de estos puntos en una base de referencia coordenada fija $R = (O, \overrightarrow{OA_1}, \ldots, \overrightarrow{OA_n})$. El punto X pertenece a la recta $B_0 \vee B_1$ cuando, y sólo cuando, existe un número λ tal que $\overrightarrow{B_0X} = \lambda \overrightarrow{B_0B_1}$, es decir, cuando

$$\xi^i - \beta_0^i = \lambda(\beta_1^i - \beta_0^i) \quad (i = 1, \ldots, n). \tag{10}$$

Determinando λ de cada una de las ecuaciones (10) e igualando los resultados, obtenemos

$$\frac{\xi^1 - \beta_0^1}{\beta_1^1 - \beta_0^1} = \frac{\xi^2 - \beta_0^2}{\beta_1^2 - \beta_0} = \ldots = \frac{\xi^n - \beta_0^n}{\beta_1^n - \beta_0^n}. \tag{11}$$

Recíprocamente, siendo $\xi^1, \ldots, \xi^n$ las coordenadas de un punto X que satisfacen las igualdades (11) e indicando por λ el valor común de las razones (11), obtenemos (10), de modo que $X \in B_0 \vee B_1$. Por consiguiente, *las ecuaciones* (11) *son las ecuaciones de la recta que pasa por dos puntos dados.*

Notemos que en las razones (11) algunos de los denominadores pueden anularse. Puesto que queremos que las relaciones (11) sean equivalentes a las condiciones (10), debemos tomar $\xi^i - \beta_0^i = 0$ siempre que $\beta_1^i - \beta_0^i = 0$.

Poniendo $\beta_1^i - \beta_0^i = \mu^i$ y $\beta_0^i = \beta^i$, podemos representar las ecuaciones (11) en la forma

$$\frac{\xi^1 - \beta^1}{\mu^1} = \frac{\xi^2 - \beta^2}{\mu^2} = \ldots = \frac{\xi^n - \beta^n}{\mu^n}. \tag{12}$$

Los números $\mu^1, \ldots, \mu^n$ se llaman *coeficientes directores* de la recta y las ecuaciones (12) se llaman *ecuaciones de la recta en coeficientes directores.*

TEOREMA 2. *Para que la recta* (12) *sea paralela a la recta*

$$\frac{\xi^1 - \gamma^1}{\nu^1} = \frac{\xi^2 - \gamma^2}{\nu^2} = \ldots = \frac{\xi^n - \gamma^n}{\nu^n} \tag{13}$$

es necesario y suficiente que sus coeficientes directores sean proporcionales:

$$\frac{\mu^1}{\nu^1} = \frac{\mu^2}{\nu^2} = \ldots = \frac{\mu^n}{\nu^n}. \tag{14}$$

Las rectas (12) *y* (13) *coinciden cuando, y sólo cuando, se cumplen las condiciones de proporcionalidad* (14) *y las condiciones*

$$\frac{\beta^1-\gamma^1}{\nu^1}=\frac{\beta^2-\gamma^2}{\nu^2}=\ldots=\frac{\beta^n-\gamma^n}{\nu^n},$$

que significan que el punto de coordenadas $\beta^1, \ldots, \beta^n$ *pertenece a la recta* (12) *se halle sobre la recta* (13).

De las ecuaciones (12) se ve que los puntos $B_0(\beta^1, \ldots, \beta^n)$ y $B_1(\beta^1+\mu^1, \ldots, \beta^n+\mu^n)$ se hallan sobre la recta que representan estas ecuaciones. Análogamente, sobre la recta (13) se hallan los puntos $C_0(\gamma^1, \ldots, \gamma^n)$ y $C_1(\gamma^1+\nu^1, \ldots, \gamma^n+\nu^n)$. Según el p. 29.3, las rectas $B_0 \vee B_1$ y $C_0 \vee C_1$ son paralelas cuando, y sólo cuando, se tenga $\overrightarrow{C_0C_1}=\lambda\overrightarrow{B_0B_1}$ para un $\lambda \in K$, es decir, cuando las coordenadas de los vectores $\overrightarrow{B_0B_1}$ y $\overrightarrow{C_0C_1}$ sean proporcionales. Puesto que las coordenadas de estos vectores son iguales, respectivamente, a $\mu^1, \ldots, \mu^n$ y $\nu^1, \ldots, \nu^n$, obtenemos de aquí (14). Como para la coincidencia de las rectas es suficiente que tengan un punto común y que sean paralelas, la segunda afirmación del teorema 2 se desprende directamente de la primera afirmación.

Además de las ecuaciones de tipo (13), una recta en un espacio afín de n dimensiones puede ser definida también mediante un sistema general de ecuaciones lineales de tipo

$$\alpha_1^i\xi^1+\ldots+\alpha_n^i\xi^n+\alpha_0^i=0 \quad (i=1, \ldots, s), \tag{15}$$

donde los rangos de la matriz principal y de la ampliada son iguales a $n-1$. ¿Cómo conociendo las ecuaciones de una recta en la forma general (15) hallar sus ecuaciones en la forma canónica (13)? Para ello es suficiente hallar las coordenadas de dos cualesquiera puntos diferentes de la recta (15) y recurrir después a la fórmula (11); pero se puede proceder también del modo siguiente. El rango del sistema (15) es, por hipótesis, igual a $n-1$. Luego, $n-1$ coordenadas variables, digamos $\xi^1, \ldots, \xi^{n-1}$, pueden ser expresadas mediante la restante coordenada variable ξ^n. Así obtenemos unas expresiones de tipo

$$\xi^i=\mu^i\xi^n+\beta^i \quad (i=1, \ldots, n-1),$$

de donde

$$\frac{\xi^1-\beta^1}{\mu^1}=\ldots=\frac{\xi^{n-1}-\beta^{n-1}}{\mu^{n-1}}=\frac{\xi^n}{1}.$$

Estas serán precisamente las ecuaciones canónicas de la recta (15).

Para concluir determinemos las condiciones de paralelismo de una recta y de un hiperplano en la forma coordenada.

TEOREMA 3. *Un hiperplano dado mediante la ecuación*

$$\alpha_1\xi^1+\ldots+\alpha_n\xi^n+\alpha_0=0 \tag{16}$$

es paralelo a una recta dada mediante las ecuaciones

$$\frac{\xi^1-\beta^1}{\mu^1}=\ldots=\frac{\xi^n-\beta^n}{\mu^n} \tag{17}$$

cuando, y sólo cuando,

$$\alpha_1\mu^1+\ldots+\alpha_n\mu^n=0. \tag{18}$$

Indicando por λ el valor común de las razones de (17), obtenemos

$$\xi^i=\mu^i\lambda+\beta^i \quad (i=1, \ldots, n). \tag{19}$$

Introduciendo en (16) estos valores en lugar de las coordenadas variables ξ^i, llegamos a la ecuación

$$(\alpha_1\mu^1+\ldots+\alpha_n\mu^n)\,\lambda=-(\alpha_1\beta^1+\ldots+\alpha_n\beta^n+\alpha_0). \tag{20}$$

Si resulta que $\alpha_1\mu^1+\ldots+\alpha_n\mu^n\neq 0$, encontraremos primero de (20) un valor único para λ y, después, de (19) unos valores, también únicos, para las coordenadas ξ^i del punto de intersección de la recta (17) y del hiperplano (16). Por consiguiente, en este caso la recta y el hiperplano son, sin duda alguna, no paralelos.

Sea $\alpha_1\mu^1+\ldots+\alpha_n\mu^n=0$. Si el segundo miembro de (20) es diferente de cero, la ecuación (20) no tiene soluciones para λ, la recta (17) y el hiperplano (16) no se cortan y, por ello, son paralelos. Por otra parte, si el segundo miembro de (20) es igual a cero, la ecuación (20) se cumple para cualesquiera valores de λ, es decir, todos los puntos de la recta se hallan sobre el hiperplano y, por consiguiente, la recta y el hiperplano son de nuevo paralelos.

30.4. Transformación de coordenadas afines. ¿Cómo varían las coordenadas $\xi^1, \ldots, \xi^n$ de un punto arbitrario X de un espacio afín $\mathfrak{A}$ de n dimensiones, si pasamos de la base de referencia coordenada fija $\boldsymbol{R}=(O, \overrightarrow{OA}_1, \ldots, \overrightarrow{OA}_n)$ a otra base de referencia $\boldsymbol{R}'=(O', \overrightarrow{O'A'_1}, \ldots, \overrightarrow{O'A'_n})$? Consideremos primero el caso en el que la base de referencia $\boldsymbol{R}'$ se obtiene de la base de referencia $\boldsymbol{R}$ mediante una traslación determinada por el vector $\overrightarrow{OO'}$. En este caso se tiene

$$\overrightarrow{O'A'_i}=\overrightarrow{OA}_i \quad (i=1, \ldots, n). \tag{1}$$

Tenemos, por hipótesis,

$$\overrightarrow{OX}=\xi^1\cdot\overrightarrow{OA}_1+\ldots+\xi^n\cdot\overrightarrow{OA}_n. \tag{2}$$

Indicando por $\alpha^1, \ldots, \alpha^n$ las coordenadas del origen nuevo O' (en la base de referencia "antigua" $\boldsymbol{R}$) y por $\xi^{1'}, \ldots, \xi^{n'}$ las coordenadas nuevas del punto X (en la base de referencia $\boldsymbol{R}'$), obtenemos

$$\overrightarrow{OO'}=\alpha^1\cdot\overrightarrow{OA}_1+\ldots+\alpha^n\cdot\overrightarrow{OA}_n, \tag{3}$$

$$\overrightarrow{O'X}=\xi^{1'}\cdot\overrightarrow{O'A'_1}+\ldots+\xi^{n'}\cdot\overrightarrow{O'A'_n}. \tag{4}$$

Sumando término a término las igualadades (3) y (4), empleando las relaciones (1) y comparando el resultado obtenido con las fórmulas (2), llegamos a las fórmulas que buscamos

$$\xi^i = \xi^{i'} + \alpha^i \quad (i = 1, \ldots, n) \tag{5}$$

que representan *la ley de transformación de las coordenadas en una traslación del origen.*

Consideremos ahora el caso en el que el origen de la base de referencia coordenada nueva coincide con el origen de la base de referencia antigua. Los sistemas de los vectores coordenados $\overline{OA_1}, \ldots, \overline{OA_n}$ y $\overline{O'A'_1}, \ldots, \overline{O'A'_n}$ representan dos bases del espacio vectorial $L(\mathfrak{A})$, de modo que los vectores coordenados nuevos $\overline{O'A'_i}$ pueden ser expresados mediante los antiguos por fórmulas de tipo

$$\overline{OA'_i} = \tau'_i \cdot \overline{OA_1} + \ldots + \tau'^n_i \cdot \overline{OA_n} \quad (i = 1, \ldots, n), \tag{6}$$

donde el determinante de la matriz $T = \|\tau'^j_i\|$ es diferente de cero (p. 8.3).

Observando que $O = O'$, introducimos en la fórmula (4) las expresiones (6) en lugar de los vectores $\overline{OA'_i}$. Comparando el resultado de la sustitución con la igualdad (2), llegamos a las relaciones

$$\xi^i = \xi^{1'}\tau'^1_i + \ldots + \xi^{n'}\tau'^n_i \quad (i = 1, \ldots, n)$$

que pueden ser representadas brevemente en la forma matricial conocida

$$(\xi^1, \ldots, \xi^n) = (\xi^{1'}, \ldots, \xi^{n'}) \cdot T. \tag{7}$$

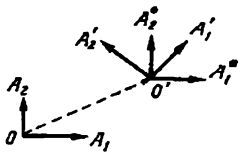


Fig. 9.

Finalmente el paso de una base de referencia $R = (O, \overline{OA_1}, \ldots, \overline{OA_n})$ a otra base cualquiera de referencia $R' = (O', \overline{O'A'_1}, \ldots, \overline{O'A'_n})$ puede ser realizado en dos pasos (véase la fig. 9 para el caso en el que $n = 2$): 1) trasladamos la base de referencia inicial en el vector $\overline{OO'}$, obteniendo así una base de referencia R^* de origen en el punto O' y 2) pasamos de la base de referencia R^* a la base de referencia R' sin cambiar el origen de coordenadas. Empleando sucesivamente las fórmulas (5) y (7), obtenemos las relaciones definitivas

$$(\xi^1, \ldots, \xi^n) = (\xi^{1'}, \ldots, \xi^{n'}) \cdot T + (\alpha^1, \ldots, \alpha^n), \tag{8}$$

donde $T = \|\tau'^j_i\|$ es la matriz del cambio de la base $\overline{OA_1}, \ldots, \overline{OA_n}$ por la base $\overline{O'A'_1}, \ldots, \overline{O'A'_n}$ (que está formada por los coeficientes de las expresiones lineales de los vectores $\overline{O'A'_i}$ en términos de los vectores $\overline{OA_1}, \ldots, \overline{OA_n}$), $(\alpha^1, \ldots, \alpha^n)$ son las coordenadas del

origen nuevo y $(\xi_1, \ldots, \xi^n)$ y $(\xi^{1'}, \ldots, \xi^{n'})$ son, respectivamente, las coordenadas antiguas y nuevas de un punto arbitrario X.

Por razones puramente formales la fila $(1, \xi^1, \ldots, \xi^n)$ ha sido llamada en el p. 30.1 fila coordenada ampliada del punto X siendo $(\xi^1, \ldots, \xi^n)$ la fila corriente de sus coordenadas. De la fórmula (8) se desprende directamente que

$$(1, \xi^1, \ldots, \xi^n) = (1, \xi^{1'}, \ldots, \xi^{n'}) \cdot T_{R'R}, \tag{9}$$

donde

$$T_{R'R} = \begin{vmatrix} 1 & \alpha^1 & \ldots & \alpha^n \\ 0 & \tau_1^1 & \ldots & \tau_n^1 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \tau_1^n & \ldots & \tau_n \end{vmatrix}. \tag{10}$$

La matriz $T_{R'R}$ se llama *matriz del cambio de la base de referencia* R *por la base de referencia* R'. De (10) resulta que el determinante de la matriz $T_{R'R}$ coincide con el determinante de la matriz $\|\tau_j^i\|$ que es la matriz del cambio de la base coordenada antigua del espacio vectorial $L(\mathfrak{A})$ por la base nueva. Recíprocamente, cualquiera que sea la matriz T_0 de tipo (10) con el determinante diferente de cero y cualquiera que sea la base de referencia R dada de antemano, las fórmulas (9) y los números $\alpha^1, \ldots, \alpha^n$ permiten determinar unívocamente una base de referencia R' tal que la matriz T_0 sea la matriz del cambio de R por R'.

De la fórmula (9) se desprende el siguiente corolario importante.

COROLARIO. *Sean* R, R' *y* R'' *tres bases arbitrarias de referencia, sea* $T_{R'R}$ *la matriz del cambio de* R *por* R' *y sea* $T_{R''R'}$ *la matriz del cambio de* R' *por* R''. *Entonces la matriz del cambio de* R *por* R'' *será igual al producto* $T_{R''R'} T_{R'R}$. *En particular, si* $T_{R'R}$ *es la matriz del cambio de la base de referencia* R *por la base de referencia* R', *la matriz del cambio de* R' *por* R *es la matriz inversa* $T_{R'R}^{-1}$.

En efecto, indicando por $(\xi^1, \ldots, \xi^n)$, $(\xi^{1'}, \ldots, \xi^{n'})$ y $(\xi^{1''}, \ldots, \xi^{n''})$ las coordenadas de un punto arbitrario X en las bases de referencia R, R' y R'', respectivamente, tenemos, según (9),

$$(1, \xi^1, \ldots, \xi^n) = (1, \xi^{1''}, \ldots, \xi^{n''}) \cdot T_{R''R}$$

y, al mismo tiempo, se tiene

$$(1, \xi^1, \ldots, \xi^n) = (1, \xi^{1'}, \ldots, \xi^{n'}) \cdot T_{R'R} = \\ = (1, \xi^{1''}, \ldots, \xi^{n''}) \cdot T_{R''R'} T_{R'R}.$$

Por consiguiente, para cualesquiera $\xi^{1''}, \ldots, \xi^{n''} \in K$ resulta

$$(1, \xi^{1''}, \ldots, \xi^{n''}) \cdot T_{R''R} = (1, \xi^{1''}, \ldots, \xi^{n''}) \cdot T_{R''R'} T_{R'R}.$$

Tomando aquí para $\xi^{1''}, \ldots, \xi^{n}$ las coordenadas en la base de referencia R'' del vértice de esta base de referencia, obtenemos

$$T_{R''R} = T_{R''R'} T_{R'R}. \tag{11}$$

Si la base de referencia R'' coincide con la base de referencia R', la matriz $T_{R''R}$ es, obviamente, la matriz unidad y la relación (11) se convierte en

$$E = T_{RR'} T_{R'R},$$

de donde resulta

$$T_{R'R} = T_{RR'}^{-1}. \tag{12}$$

Hasta el momento nos hemos interesado sólo por la ley de variación de las coordenadas de los puntos. Preguntémonos ¿cómo varían las ecuaciones de los conjuntos de puntos cuando se pasa de una base de referencia coordenada a otra?

Sea dada una ecuación

$$f(\xi^1, \ldots, \xi^n) = 0 \tag{13}$$

de un conjunto $\mathfrak{M}$ en una base de referencia R. Esto significa, por definición, que un punto arbitrario X pertenece a $\mathfrak{M}$ cuando, y sólo cuando, sus coordenadas $\xi^1, \ldots, \xi^n$ satisfacen la relación (13). Al pasar a una base de referencia coordenada nueva R', tenemos según (8)

$$\xi^i = \sum_{j=1}^{n} \xi^{j'} \tau_j^i + \alpha^i \quad (i = 1, \ldots, n).$$

Introduciendo en (13) estas expresiones para $\xi^1, \ldots, \xi^n$, obtenemos

$$f\left(\sum \xi^{j'} \tau_j^1 + \alpha^1, \ldots, \sum \xi^{j'} \tau_j^n + \alpha^n\right) = 0. \tag{14}$$

Puesto que las coordenadas nuevas $\xi^{1'}, \ldots, \xi^{n'}$ de un punto arbitrario X satisfacen la relación (14) cuando, y sólo cuando, $X \in \mathfrak{M}$, resulta que (14) es la ecuación que buscamos del conjunto $\mathfrak{M}$ en la base de referencia coordenada nueva. El primer miembro de la igualdad (14) es una función $g(\xi^{1'}, \ldots, \xi^{n'})$ de las variables $\xi^{1'}, \ldots, \xi^{n'}$ que se distingue, en general, por su forma de la función inicial f. Sin embargo, si f es un polinomio de un grado s en las variables $\xi^1, \ldots, \xi^n$, se ve de (14) que g también será un polinomio de grado s en las variables $\xi^{1'}, \ldots, \xi^{n'}$.

Igual que en el caso de los espacios vectoriales (véase el p. 5.1), el problema sobre la transformación de las coordenadas afines está ligado estrechamente al problema de la determinación de todos los automorfismos de un espacio afín $\mathfrak{A}$ sobre un cuerpo conmutativo fijo K. Sea n la dimensión del espacio $\mathfrak{A}$ y sea $\mathcal{A}$ un automorfismo de $\mathfrak{A}$ sobre K (véase el p. 29.1). Tomemos en $\mathfrak{A}$ una

base de referencia coordenada $R=(O, \overline{OA}_1, \ldots, \overline{OA}_n)$. Sus vértices $O, A_1, \ldots, A_n$ forman en $\mathfrak{A}$ un sistema linealmente independiente de puntos. Com la relación de independencia lineal se conserva en los automorfismos, los puntos $O\mathcal{A}, A_1\mathcal{A}, \ldots, A_n\mathcal{A}$ pueden ser considerados como los vértices de una base de referencia coordenada nueva $R\mathcal{A}=(O\mathcal{A}, \overline{O\mathcal{A}A_1\mathcal{A}}, \ldots, \overline{O\mathcal{A}A_n\mathcal{A}})$. Indiquemos por $(X)^1_R, \ldots, (X)^n_R$ las coordenadas de un punto arbitrario X calculadas en la base de referencia R y sea

$$[X]_R=(1, (X)^1_R, \ldots, (X)^n_R)$$

la fila ampliada de coordenadas del punto X en la base de referencia R. La fórmula (2) toma entonces la forma siguiente:

$$\overline{OX}=(X)^1_R\cdot\overline{OA}_1+\ldots+(X)^n_R\cdot\overline{OA}_n. \tag{15}$$

Como los isomorfismos conservan las relaciones de dependencia lineal entre vectores, obtenemos de (15)

$$\overline{O\mathcal{A}X\mathcal{A}}=(X)^1_R\cdot\overline{O\mathcal{A}A_1\mathcal{A}}+\ldots+(X)^n_R\cdot\overline{O\mathcal{A}A_n\mathcal{A}}. \tag{16}$$

En otras palabras, las coordenadas de $X\mathcal{A}$ en la base de referencia $R\mathcal{A}$ coinciden con las coordenadas de X en la base de referencia R, es decir, en notación abreviada

$$[X\mathcal{A}]_{R\mathcal{A}}=[X]_R.$$

Aplicando la fórmula (9), obtenemos

$$[X]_R=[X\mathcal{A}]_{R\mathcal{A}}=[X\mathcal{A}]\cdot T_{R\mathcal{A}R}$$

y por esto, debido a (12), resulta

$$[X\mathcal{A}]_R=[X]\,T_{RR\mathcal{A}}. \tag{17}$$

La matriz $T_{RR\mathcal{A}}$ se llama *matriz del automorfismo* $\mathcal{A}$ en la base de referencia R y se indica por $[\mathcal{A}]_R$ o simplemente por $[\mathcal{A}]$, si la base de referencia R se conoce de antemano. Por esta razón la relación (17) puede ser representada en la forma definitiva

$$[X\mathcal{A}]=[X]\,[\mathcal{A}]. \tag{18}$$

Sean $\mathcal{A}$ y $\mathcal{B}$ dos automorfismos arbitrarios dados del espacio $\mathfrak{A}$. Aplicando la fórmula (18), obtenemos

$$[X(\mathcal{A}\mathcal{B})]=[X]\,[\mathcal{A}\mathcal{B}],$$
$$[(X\mathcal{A})\mathcal{B})]=[X\mathcal{A}][\mathcal{B}]=[X]\,[\mathcal{A}]\,[\mathcal{B}],$$

de donde resulta

$$[X]\,[\mathcal{A}\mathcal{B}]=[X]([\mathcal{A}]\,[\mathcal{B}]).$$

Puesto que esta igualdad debe cumplirse para cualquier punto X, tenemos

$$[\mathcal{AB}] = [\mathcal{A}]\,[\mathcal{B}]$$

y, por consiguiente,

$$[\mathcal{A}^{-1}] = [\mathcal{A}]^{-1}.$$

Consideremos un ejemplo. Sea $\mathcal{A}$ una traslación de $\mathfrak{A}$ de vector $\overline{OO'}$ de coordenadas $\alpha^1, \ldots, \alpha^n$ en la base de referencia $R = (O, \overline{OA_1}, \ldots, \overline{OA_n})$. De las fórmulas (9) y (5) obtenemos entonces

$$[\mathcal{A}] = T_{RR\mathcal{A}} = \begin{bmatrix} 1 & -\alpha^1 & -\alpha^2 & \ldots & -\alpha^n \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix}.$$

En cambio, si el automorfismo $\mathcal{A}$ deja inmóvil el origen de la base de referencia coordenada R, la matriz $[\mathcal{A}]$ se descompone en 1 y en la matriz $\|\tau_j^i\|$ del cambio en el espacio vectorial $L(\mathfrak{A})$.

Ejemplos y problemas

1. Hállense las condiciones necesarias y suficientes de intersección de dos rectas dadas por ecuaciones en coeficientes directores.

2. Hállense las ecuaciones paramétricas del plano dado por el sistema de ecuaciones

$$\left.\begin{aligned} x_1 + x_2 - 2x_3 + 3x_4 &= 1, \\ x_1 + 2x_3 - x_3 + 2x_4 &= 3, \\ x_1 - x_2 - 4x_3 - 5x_4 &= -3. \end{aligned}\right\}$$

3. Demuéstrese que todo plano $\mathfrak{M}$ de un espacio afín es el mismo un espacio afín cuya dimensión es igual a la dimensión de $\mathfrak{M}$.

4. Demuéstrese que un plano $\mathfrak{M}$ de un espacio afín diferente de un punto es paralelo a cualquier plano que no corte el primero cuando, y sólo cuando, $\mathfrak{M}$ es un hiperplano.

§ 31. Cuerpos convexos

Hasta el momento no hemos considerado conceptos geométricos tan esenciales como son los conceptos de la relación "hallarse entre", de semiplano y de convexidad. La razón de esto estriba en que en los espacios afines sobre un cuerpo conmutativo arbitrario estos conceptos no pueden ser introducidos de un modo natural. Para tratarlos es necesario limitar la clase de cuerpos conmutativos y pasar al estudio de los espacios afines sobre cuerpos conmutativos ordenados, por ejemplo, sobre el cuerpo de los números reales. Las

propiedades principales de estos espacios constituyen precisamente el tema que se estudia en este parágrafo.

31.1. Rayos. Recordemos que un cuerpo conmutativo K se llama *ordenado*, si para los elementos de K, además de las operaciones de adición y de multiplicación, se introduce también una relación de orden $\leqslant$ sometida a dos condiciones:

$$\alpha \leqslant \beta \Rightarrow \alpha + \gamma \leqslant \beta + \gamma, \tag{1}$$

$$0 \leqslant \alpha \quad \text{y} \quad 0 \leqslant \beta \Rightarrow 0 \leqslant \alpha\beta \tag{2}$$

$$(\alpha,\ \beta,\ \gamma \in K).$$

El ejemplo más importante de un cuerpo conmutativo ordenado es el cuerpo de los números reales que es el que debe tenerse en primer orden en cuenta en todos los razonamientos ulteriores de este parágrafo y del siguiente.

No está de más observar que de (1) y (2) se desprenden directamente las siguientes propiedades:

Para cualquier elemento α de un cuerpo conmutativo ordenado se tiene

$$0 \leqslant \alpha^2 \quad \text{y} \quad -\alpha^2 \leqslant 0 \tag{3}$$

y, en particular, $0 \leqslant 1$ y $-1 \leqslant 0$.

Para unos elementos arbitrarios α, β y γ de un cuerpo conmutativo ordenado se tiene

$$\alpha \leqslant \beta \Rightarrow -\beta \leqslant -\alpha, \tag{4}$$

$$0 \leqslant \alpha \quad \text{y} \quad \beta \leqslant \gamma \Rightarrow \alpha\beta \leqslant \alpha\gamma. \tag{5}$$

Se toma por definición que la relación $\alpha < \beta$ equivale a la conjunción $\alpha \leqslant \beta$ y $\alpha \neq \beta$ y que las relaciones $\alpha \geqslant \beta$ y $\alpha > \beta$ significan lo mismo que las relaciones $\beta \leqslant \alpha$ y $\beta < \alpha$. El elemento no negativo de los elementos α y $-\alpha$ se llama *valor absoluto de α* y se indica por $|\alpha|$. De las relaciones de (1) a (5) se deduce fácilmente que

$$|-\alpha| = |\alpha|, \quad |\alpha + \beta| \leqslant |\alpha| + |\beta| \quad \text{y} \quad |\alpha\beta| = |\alpha| \cdot |\beta|.$$

Consideremos ahora un espacio afín $\mathfrak{A}$ sobre un cuerpo conmutativo *ordenado* K. Se dice que en el espacio afín $\mathfrak{A}$ el punto X *se halla entre* los puntos A y B, si

$$\overrightarrow{AX} = \lambda \cdot \overrightarrow{AB} \quad \text{y} \quad 0 \leqslant \lambda \leqslant 1 \quad (\lambda \in K). \tag{6}$$

Está claro que si X se halla entre A y B, los puntos A, X y B se hallan sobre una misma recta. Además, la relación «hallarse entre» es simétrica: *si X se halla entre A y B, X se halla entre B y A.* Efectivamente, de (6) obtenemos

$$\overrightarrow{AB} = \overrightarrow{AX} + \overrightarrow{XB} = \lambda \cdot \overrightarrow{AB} + \overrightarrow{XB},$$

y por esto

$$\overrightarrow{BX} = (1-\lambda) \cdot \overrightarrow{BA}, \quad 0 \leqslant 1-\lambda \leqslant 1.$$

Sean A, B y C unos puntos, diferentes dos a dos, que se hallan sobre una misma recta. Para un valor adecuado $\lambda \in K$ tenemos entonces $\overline{AC} = \lambda \cdot \overline{AB}$. El número λ satisface a una de las condiciones, y sólo a una de ellas: a) $\lambda < 0$; b) $0 \leqslant \lambda \leqslant 1$ y c) $\lambda > 1$. Es fácil comprobar que en el caso a) el punto A se halla entre B y C, que en el caso c) el punto B se halla entre A y C y que en el caso b) el punto C se halla, por definición, entre A y B.

Por consiguiente, *de cualesquiera tres puntos que se encuentran sobre una misma recta uno, y sólo uno, se halla entre los otros dos.*

Análogamente se comprueba también la segunda propiedad principal de la relación «hallarse entre»: *si el punto X se halla entre los puntos A y B y el punto Y se halla entre A y X, el punto Y se halla entre A y B.*

Empleando el concepto «entre», podemos introducir en toda recta dos relaciones naturales de orden llamadas también direcciones de la recta. Tomemos sobre la recta dada $\mathfrak{M}$ dos puntos distintos cualesquiera A y B e introduzcamos para los puntos de $\mathfrak{M}$ una relación binaria $\leqslant_{AB}$ que depende de A y de B tomando, por definición, que $X \leqslant_{AB} Y$ es verídica si

$$\overline{AX} = \lambda \cdot \overline{AB}, \quad \overline{AY} = \mu \cdot \overline{AB} \quad \text{y} \quad \lambda \leqslant \mu \ (\lambda, \mu \in K), \tag{7}$$

es decir, ordenando los puntos de la recta $\mathfrak{M}$ según el orden en el que se encuentran sus coordenadas en la base de referencia $(A, \overline{AB})$.

Está claro que los órdenes $\leqslant_{AB}$ y $\leqslant_{BA}$ *son duales* es decir, que

$$X \leqslant_{AB} Y \Longleftrightarrow Y \leqslant_{BA} X.$$

Por otra parte, el orden $\leqslant_{PQ}$ definido sobre $\mathfrak{M}$ mediante otro par cualquiera de puntos P, Q coincide con el orden $\leqslant_{AB}$, si $P \leqslant_{AB} Q$, y coincide con el orden $\leqslant_{BA}$, si $Q \leqslant_{AB} P$. Por consiguiente, *entre los órdenes de tipo $\leqslant_{PQ}$ sobre una recta $\mathfrak{M}$ existen solamente dos órdenes diferentes.* Estos órdenes se llaman *órdenes* o *direcciones naturales* de la recta. Los automorfismos del espacio $\mathfrak{A}$ sobre K conservan la relación «hallarse entre» y también conservan los órdenes naturales. Sin embargo, gracias a la existencia sobre una recta de dos órdenes naturales, los automorfismos del espacio $\mathfrak{A}$ pueden transformar un orden natural sobre la recta $\mathfrak{M}$ en el otro orden natural sobre la misma recta. Por esto, es invariante el par de relaciones duales de órdenes y no la relación de orden.

TEOREMA. *Si una funcional lineal $f(X)$ no es constante sobre una recta $\mathfrak{M}$, la aplicación $f(X) \longrightarrow X$ $(X \in \mathfrak{M})$ es una aplicación biyectiva de K sobre $\mathfrak{M}$ en la que el orden $\leqslant$, definido sobre el cuerpo conmutativo K, se transforma en uno de los dos órdenes naturales definidos sobre $\mathfrak{M}$.*

Fijemos sobre la recta unos puntos O y A tales que $f(O) \neq f(A)$ y sea

$$\overline{OX} = \lambda \cdot \overline{OA} \quad (X \in \mathfrak{M} \quad \text{y} \quad \lambda \in K).$$

La aplicación $\lambda \to X$, es, debido a (7), una aplicación biyectiva de K sobre $\mathfrak{M}$ que transforma el orden dado sobre K en un orden natural sobre $\mathfrak{M}$. De la linealidad de f se deduce que

$$f(X) = \lambda(f(A) - f(O)) + f(O) = \alpha\lambda + \beta.$$

Puesto que $\alpha \neq 0$, la aplicación $f(X) \to \lambda$ es una aplicación biyectiva de K sobre K que o bien conserva el orden o bien lo invierte. La aplicación $f(X) \to X$, por ser la composición de las aplicaciones $f(X) \to \lambda$ y $\lambda \to X$, también es biyectiva y transforma el orden sobre K en un orden natural sobre $\mathfrak{M}$.

Cualesquiera que sean dos puntos A y B de un espacio $\mathfrak{A}$ se llama *segmento* $[A, B]$ el conjunto de todos los puntos que se hallan entre A y B. Los puntos A y B se llaman *extremos del segmento* $[A, B]$. De la definición de la relación «hallarse entre» se deduce que los extremos del segmento pertenecen al segmento y de la simetría de esta relación se deduce que $[A, B] = [B, A]$. Está claro que el segmento $[A, A]$ está formado sólo por el punto A.

Se dice que un número λ se halla entre los números α y β ($\alpha, \beta, \lambda \in K$), si $\alpha \leqslant \lambda \leqslant \beta$ o $\beta \leqslant \lambda \leqslant \alpha$. El conjunto de los números que se hallan entre α y β se llama segmento numérico y se indica por $[\alpha, \beta]$. Del teorema anterior obtenemos el siguiente corolario:

COROLARIO. *El conjunto de los valores que toma una funcional lineal $f(X)$ sobre un segmento $[A, B]$ es un segmento numérico $[f(A), f(B)]$. El conjunto de los valores que toma f sobre la recta $\mathfrak{M}$ o bien coincide con todo el cuerpo conmutativo K (si f no es constante sobre $\mathfrak{M}$) o bien está formado por un solo número (si f es constante sobre $\mathfrak{M}$).*

Introduzcamos ahora el concepto de semirrecta o de rayo. Se dice que el punto X se encuentra al mismo lado del punto O que el punto A, si X se halla entre O y A o si A se halla entre O y X.

Un conjunto de puntos $\mathfrak{M}_0$ del espacio $\mathfrak{A}$ se llama *rayo*, si en $\mathfrak{M}_0$ existen unos puntos distintos O y A tales que $\mathfrak{M}_0$ está formado por todos los puntos X que se encuentran al mismo lado del punto O que el punto A. El punto O se llama vértice del rayo que será indicado ahora por $\mathfrak{M}_{OA}$.

De esta definición se deduce que todos los puntos del rayo $\mathfrak{M}_{OA}$ pertenecen a la recta $O \vee A$. Si tomamos como base de referencia coordenada sobre la recta $O \vee A$ la base de referencia $(O, \overline{OA})$, a todo $\lambda \in K$ le corresponderá un punto X de coordenada lineal λ tal que $\overline{OX} = \lambda\overline{OA}$. Está claro que el rayo $\mathfrak{M}_{OA}$ consta de todos los puntos de la recta $O \vee A$ que tienen coordenada no negativa. En particular, de aquí se ve que todo rayo tiene solamente un vértice O y que si sobre el rayo $\mathfrak{M}_{OA}$ se toma un punto cualquiera B diferente de O, los rayos $\mathfrak{M}_{OA}$ y $\mathfrak{M}_{OB}$ coinciden. Es fácil comprobar también que al tomar en una recta un punto arbitrario O, la recta se descompone exactamente en dos rayos diferentes de vértice O.

Supongamos que una funcional lineal $f(X)$ no es constante sobre una recta $\mathfrak{M}$. En virtud del teorema 1, sobre la recta $\mathfrak{M}$ existe un punto O, sólo uno, en el que la funcional f se anula. Puesto que la aplicación $f(X) \to X$ transforma el orden de K en un orden natural de $\mathfrak{M}$, los conjuntos $\mathfrak{M}_0$ y $\mathfrak{M}_1$ de puntos de $\mathfrak{M}$ con $\lambda \leqslant 0$ y $\lambda \geqslant 0$, respectivamente, son precisamente aquellos rayos en los que el punto O divide a la recta $\mathfrak{M}$.

31.2. Semiespacios. El análogo del concepto de rayo en el caso de planos multidimensionales de un espacio $\mathfrak{A}$ es el concepto de semiplano y, en particular, el concepto de semiespacio que pasamos ahora a definir.

Sea de nuevo $\mathfrak{A}$ un espacio afín sobre un cuerpo conmutativo ordenado K y sea $\mathfrak{P}$ un hiperplano arbitrario de $\mathfrak{A}$. Si $\mathfrak{P}$ contiene unos puntos X e Y, $\mathfrak{P}$ contiene también todos los puntos de la recta $X \vee Y$. Por esto, cualquier segmento $[A, B]$ de $\mathfrak{A}$ o bien está contenido íntegramente en $\mathfrak{P}$ o bien contiene no más de un punto de $\mathfrak{P}$. Se dice que los puntos A y B *se hallan a distintos lados* del hiperplano $\mathfrak{P}$, si ambos no pertenecen a $\mathfrak{P}$, pero el segmento $[A, B]$ contiene un punto de $\mathfrak{P}$. En todos los demás casos se dice que los puntos A y B *se hallan a un mismo lado* de $\mathfrak{P}$ lo que se expresa simbólicamente así: $A = B(\mathfrak{P})$. En particular, $A = A(\mathfrak{P})$ y de $A \equiv B(\mathfrak{P})$ se deduce que $B \equiv A(\mathfrak{P})$ para cualesquiera A y B. Además, si $A \in \mathfrak{P}$, se tiene $A = B(\mathfrak{P})$ para cualquier punto B.

TEOREMA 1. *Sea $f(X)$ una funcional lineal no constante sobre el espacio $\mathfrak{A}$ y sea $\mathfrak{P}$ un hiperplano formado por los puntos X tales que $f(X) = 0$. Entonces se tiene para cualesquiera A y B*

$$A = B(\mathfrak{P}) \Longleftrightarrow f(A) \cdot f(B) \geqslant 0. \qquad (1)$$

Supongamos que A y B se hallan a distintos lados de $\mathfrak{P}$. El conjunto de los valores que toma $f(X)$ en los puntos del segmento $[A, B]$ es el segmento numérico $[f(A), f(B)]$. Tenemos, por hipótesis, $f(A) \neq 0$, $f(B) \neq 0$ y $0 \in [f(A), f(B)]$. Por consiguiente $f(A) \cdot f(B) < 0$. Recíprocamente, supongamos que para unos puntos A y B se tiene $f(A) \cdot f(B) < 0$. Entonces $0 \in [f(A), f(B)]$, es decir, para un punto $X \in [A, B]$ se tiene $f(X) = 0$, de donde $X \in \mathfrak{P}$ y por ello los puntos A y B se hallan a distintos lados de $\mathfrak{P}$.

COROLARIO. *Si el punto A no pertenece al hiperplano $\mathfrak{P}$ y $A \equiv B(\mathfrak{P})$ y $A \equiv C(\mathfrak{P})$, se tiene $B = C(\mathfrak{P})$.*

Efectivamente, si los puntos A, B y C satisfacen las exigencias indicadas, tenemos según el teorema 1

$$f(A) \neq 0, \quad f(A) f(B) \geqslant 0 \quad \text{y} \quad f(A) f(C) \geqslant 0,$$

de donde $(f(A))^2 f(B) f(C) \geqslant 0$ y, por consiguiente, $f(B) f(C) \geqslant 0$.

Introduzcamos ahora el concepto principal: cualesquiera que sean un hiperplano $\mathfrak{P}$ y un punto arbitrario $A \notin \mathfrak{P}$, el conjunto de los puntos que se hallan a un mismo lado de $\mathfrak{P}$ que el punto A

se llama *semiespacio definido por el hiperplano* $\mathfrak{P}$ *y por el punto* A. Convengamos en indicar provisionalmente este semiespacio por $\mathfrak{A}_{\mathfrak{P}A}$.

Se ve de la definición que $\mathfrak{A}_{\mathfrak{P}A}$ contiene indudablemente el punto A y el hiperplano $\mathfrak{P}$. El conjunto que se obtiene de $\mathfrak{A}_{\mathfrak{P}A}$ omitiendo el hiperplano $\mathfrak{P}$ se llama *semiespacio abierto* definido por el hiperplano $\mathfrak{P}$ y por el punto A.

TEOREMA 2 *Si los puntos* A *y* B *no pertenecen al hiperplano* $\mathfrak{P}$ *y el segmento* $[A, B]$ *no contiene puntos de* $\mathfrak{P}$, *los semiespacios* $\mathfrak{A}_{\mathfrak{P}A}$ *y* $\mathfrak{A}_{\mathfrak{P}B}$ *coinciden. Si los puntos* A *y* C *se hallan a distintos lados de* $\mathfrak{P}$, *se tiene* $\mathfrak{A}_{\mathfrak{P}A} \neq \mathfrak{A}_{\mathfrak{P}C}$ *y cualquier semiespacio* $\mathfrak{A}_{\mathfrak{P}D}$ *coincide con* $\mathfrak{A}_{\mathfrak{P}A}$ *o con* $\mathfrak{A}_{\mathfrak{P}C}$.

Consideremos una funcional lineal no constante $f(X)$ que se anula sobre $\mathfrak{P}$. Puesto que el segmento $[A, B]$ no contiene puntos de $\mathfrak{P}$, los números $f(A)$ y $f(B)$ son de un mismo signo debido a la fórmula (1). Análogamente se comprueba que los números $f(A)$ y $f(C)$ tienen signos diferentes. Del teorema 1 resulta que para cualquier punto $D \notin \mathfrak{P}$ siendo $f(D) > 0$ el conjunto $\mathfrak{A}_{\mathfrak{P}D}$ coincide con la colección $\mathfrak{A}^+$ de los puntos X tales que

$$f(X) \geqslant 0 \tag{2}$$

y siendo $f(D) < 0$ el conjunto $\mathfrak{A}_{\mathfrak{P}D}$ coincide con la colección $\mathfrak{A}^-$ de los puntos X tales que

$$f(X) \leqslant 0. \tag{3}$$

Por consiguiente, un semiespacio arbitrario $\mathfrak{A}_{\mathfrak{P}D}$ coincide con $\mathfrak{A}^+$ o con $\mathfrak{A}^-$. Si $f(A) > 0$, de las relaciones mencionadas obtenemos

$$\mathfrak{A}_{\mathfrak{P}A} = \mathfrak{A}_{\mathfrak{P}B} = \mathfrak{A}^+ \quad \text{y} \quad \mathfrak{A}_{\mathfrak{P}C} = \mathfrak{A}^-;$$

en cambio, si $f(A) < 0$, se tiene

$$\mathfrak{A}_{\mathfrak{P}A} = \mathfrak{A}_{\mathfrak{P}B} = \mathfrak{A}^- \quad \text{y} \quad \mathfrak{A}_{\mathfrak{P}C} = \mathfrak{A}^+.$$

COROLARIO 1. *Para toda funcional lineal no constante* $f(X)$ *el conjunto de los puntos* X *que satisfacen la desigualdad*

$$f(X) \geqslant 0$$

y el conjunto de los puntos X *que satisfacen la desigualdad*

$$f(X) \leqslant 0$$

representan los dos semiespacios definidos por el hiperplano

$$f(X) = 0. \tag{4}$$

Los conjuntos de los puntos X *definidos por la desigualdad estricta*

$$f(X) > 0 \tag{5}$$

y, respectivamente, por la desigualdad estricta

$$f(X) < 0 \tag{6}$$

representan los semiespacios abiertos en los que el hiperplano (4) *divide al espacio afín dado* $\mathfrak{A}$.

Efectivamente, la ecuación (4) determina, según el p. 29.4, un hiperplano $\mathfrak{P}$ que define los semiespacios (2) y (3). Los conjuntos de los puntos definidos por las desigualdades estrictas (5) y (6) se obtienen de los semiespacios (2) y (3) omitiendo las soluciones de la ecuación (4) y son, por consiguiente, semiespacios abiertos.

COROLARIO 2. *Supongamos que en un espacio* $\mathfrak{A}$ *de dimensión finita* n *se ha escogido una base de referencia coordenada cualquiera. El conjunto de los puntos* X *cuyas coordenadas* $\xi^1, \ldots, \xi^n$ *satisfacen una desigualdad determinada*

$$\alpha_1\xi^1 + \ldots + \alpha_n\xi^n \geqslant \alpha_0$$

en la que al menos uno de los coeficientes principales $\alpha_1, \ldots, \alpha_n$ *es diferente de cero, representa entonces un semiespacio definido por el hiperplano*

$$\alpha_1\xi^1 + \ldots + \alpha_n\xi^n = \alpha_0. \tag{7}$$

El conjunto de los puntos X *cuyas coordenadas satisfacen la desigualdad estricta*

$$\alpha_1\xi^1 + \ldots + \alpha_n\xi^n > \alpha_0$$

representa un semiespacio abierto definido por el hiperplano (7).

La funcional definida mediante la fórmula

$$f(X) = \alpha_1\xi^1 + \ldots + \alpha_n\xi^n - \alpha_0$$

es, por lo visto en el p. 30.1, lineal y no constante sobre $\mathfrak{A}$. Aplicando a f el corolario 1, obtenemos las afirmaciones requeridas.

Hemos considerado hasta el momento los casos, en cierto sentido, extremos: rayos o semirrectas y semiespacios. No obstante, es fácil definir también el concepto de semiplano para un plano cualquiera $\mathfrak{M}$ contenido en el espacio $\mathfrak{A}$. En efecto, sabemos que $\mathfrak{M}$ puede ser considerado como un subespacio sobre el mismo cuerpo conmutativo K sobre el cual está definido el espacio $\mathfrak{A}$. Aplicando lo expuesto anteriormente al espacio afín $\mathfrak{M}$ sobre K, obtenemos el concepto de semiespacio del espacio $\mathfrak{M}$. Estos semiespacios se llaman *semiplanos* del plano $\mathfrak{M}$.

TEOREMA 3. *Supongamos que el plano* $\mathfrak{M}$ *y el hiperplano* $\mathfrak{P}$ *no son paralelos y sea* A *un punto de* $\mathfrak{M}$ *que no pertenece a* $\mathfrak{P}$. *La intersección del semiespacio* $\mathfrak{A}_{\mathfrak{P}A}$ *y del plano* $\mathfrak{M}$ *será entonces un semiplano en* $\mathfrak{M}$ *definido por el punto* A *y el hiperplano* $\mathfrak{M} \cap \mathfrak{P}$ *del espacio* $\mathfrak{M}$.

Todo semiplano de $\mathfrak{M}$ *es la intersección de* $\mathfrak{M}$ *con un semiespacio adecuado* $\mathfrak{A}_{\mathfrak{P}A}$ *del espacio* $\mathfrak{A}$.

Sea $f(X)$ una funcional lineal no constante que se anula sobre $\mathfrak{P}$ y sea $f(A) > 0$. El semiespacio $\mathfrak{A}_{\mathfrak{P}A}$ es entonces el conjunto de las soluciones de la desigualdad $f(X) \geqslant 0$ y la intersección $\mathfrak{M} \cap \mathfrak{A}_{\mathfrak{P}A}$

es el conjunto de los puntos $Y \in \mathfrak{M}$ que satisfacen la desigualdad $f(Y) \geqslant 0$. Puesto que la contracción de f sobre $\mathfrak{M}$ es una funcional lineal sobre $\mathfrak{M}$, resulta, según el corolario 1, que el conjunto $\mathfrak{M} \cap \mathfrak{A}_{\mathfrak{P}A}$ es un semiplano en $\mathfrak{M}$. Análogamente se demuestra la segunda afirmación del teorema 3.

Volviendo de nuevo a los semiespacios del espacio $\mathfrak{A}$ plantéemonos el problema: *¿de qué modo pueden situarse con respecto uno al otro dos semiespacios* $\mathfrak{A}_{\mathfrak{P}A}$ *y* $\mathfrak{A}_{\mathfrak{Q}B}$*?*

Consideremos por separado cada uno de los tres casos que pueden darse aquí:

1) Los hiperplanos $\mathfrak{P}$ y $\mathfrak{Q}$ coinciden. Entonces los semiespacios $\mathfrak{A}_{\mathfrak{P}A}$ y $\mathfrak{A}_{\mathfrak{Q}B}$ o bien coinciden o bien la unión de ellos es todo el espacio $\mathfrak{A}$ y la intersección es el hiperplano $\mathfrak{P}$.

2) Los hiperplanos $\mathfrak{P}$ y $\mathfrak{Q}$ no se cortan y, por consiguiente, son paralelos. Aquí pueden darse los siguientes subcasos:

a) $\mathfrak{P} \notin \mathfrak{A}_{\mathfrak{Q}B}$ y $\mathfrak{Q} \notin \mathfrak{A}_{\mathfrak{P}A}$; entonces $\mathfrak{A}_{\mathfrak{P}A} \cap \mathfrak{A}_{\mathfrak{Q}B} = \varnothing$.

b) $\mathfrak{P} \in \mathfrak{A}_{\mathfrak{Q}B}$ y $\mathfrak{Q} \notin \mathfrak{A}_{\mathfrak{P}A}$; entonces $\mathfrak{A}_{\mathfrak{Q}B} \supset \mathfrak{A}_{\mathfrak{P}A}$.

c) $\mathfrak{P} \notin \mathfrak{A}_{\mathfrak{Q}B}$ y $\mathfrak{Q} \in \mathfrak{A}_{\mathfrak{P}A}$; entonces $\mathfrak{A}_{\mathfrak{P}A} \supset \mathfrak{A}_{\mathfrak{Q}B}$.

d) $\mathfrak{P} \in \mathfrak{A}_{\mathfrak{Q}B}$ y $\mathfrak{Q} \in \mathfrak{A}_{\mathfrak{P}A}$; entonces $\mathfrak{A}_{\mathfrak{P}A} \cap \mathfrak{A}_{\mathfrak{Q}B} \neq \varnothing$.

Los conjuntos del último tipo se llaman a veces *hipercapas*. La intersección de cualquier recta con una hipercapa es, como puede verse fácilmente, vacía o coincide con toda la recta o es un segmento de la recta.

3) Los hiperplanos $\mathfrak{P}$ y $\mathfrak{Q}$ se cortan. El conjunto $\mathfrak{P} \cap \mathfrak{Q}$ es, según el p. 29.2, un plano de codimensión 2. En este caso la intersección de los semiespacios $\mathfrak{A}_{\mathfrak{P}A} \cap \mathfrak{A}_{\mathfrak{Q}B}$ se llama *hiperángulo* de arista $\mathfrak{P} \cap \mathfrak{Q}$.

31.3. Conjuntos convexos. Un conjunto $\mathfrak{S}$ de puntos de un espacio afín $\mathfrak{A}$ sobre un cuerpo conmutativo ordenado K se llama *convexo*, si

$$X \in \mathfrak{S} \text{ e } Y \in \mathfrak{S} \Rightarrow [X, Y] \in \mathfrak{S}. \tag{1}$$

De aquí se deduce que los planos del espacio $\mathfrak{A}$, así como los segmentos, rayos, semiplanos y semiplanos abiertos de $\mathfrak{A}$ son conjuntos convexos.

De la definición (1) se ve directamente que *la intersección de cualquier familia de conjuntos convexos es un conjunto convexo.* En particular, la intersección de cualquier familia de semiespacios es un conjunto convexo.

La intersección de todos los conjuntos convexos que contienen un conjunto fijo de puntos $\mathfrak{M}$ se llama *adherencia convexa del conjunto* $\mathfrak{M}$ y se indica por $\mathrm{Conv}\mathfrak{M}$. De la observación hecha anteriormente se desprende que la adherencia convexa de un conjunto cualquiera $\mathfrak{M}$ es el menor conjunto convexo que contiene a $\mathfrak{M}$. Está claro que $\mathfrak{M}$ es convexo cuando, y sólo cuando, $\mathrm{Conv}\mathfrak{M} = \mathfrak{M}$.

También es evidente que de $\mathfrak{M} \subseteq \mathfrak{N}$ resulta $\operatorname{Conv} \mathfrak{M} \subseteq \operatorname{Conv} \mathfrak{N}$. Todo punto $X \in \mathfrak{A}$ constituye por sí mismo un conjunto convexo y, por consiguiente, $\operatorname{Conv} X = X$. Sin embargo, ya para dos puntos X e Y tenemos, como puede verse fácilmente, la fórmula

$$\operatorname{Conv}\{X, Y\} = [X, Y]. \tag{2}$$

El teorema que sigue ofrece la expresión para la adherencia convexa de un conjunto cualquiera.

TEOREMA 1. *Escojamos en el espacio $\mathfrak{A}$ un punto O. La adherencia convexa* $\operatorname{Conv} \mathfrak{M}$ *de un conjunto arbitrario de puntos $\mathfrak{M}$ es el conjunto de todos los puntos $X \in \mathfrak{A}$ tales que para cada uno de ellos existe en $\mathfrak{M}$ un sistema finito de puntos $M_1, \ldots, M_s$ ligados a X por la relación*

$$\overline{OX} = \lambda_1 \cdot \overline{OM_1} + \ldots + \lambda_s \cdot \overline{OM_s}, \tag{3}$$

donde $\lambda_1, \ldots, \lambda_s$ son unos números adecuados de K sometidos a las condiciones

$$\lambda_1 \geqslant 0, \ \ldots, \ \lambda_s \geqslant 0, \tag{4}$$

$$\lambda_1 + \ldots + \lambda_s = 1. \tag{5}$$

Indiquemos por $\mathfrak{M}_c$ el conjunto de los puntos X que cumplen las exigencias (3), (4) y (5) y demostremos que $\mathfrak{M}_c$ es convexo. Sean $A, B \in \mathfrak{M}_c$, de modo que para convenientes $M_i \in \mathfrak{M}$ y $\alpha_i, \beta_i \in K$ se tenga

$$\overline{OA} = \alpha_1 \cdot \overline{OM_1} + \ldots + \alpha_s \cdot \overline{OM_s},$$

$$\overline{OB} = \beta_1 \cdot \overline{OM_1} + \ldots + \beta_s \cdot \overline{OM_s},$$

donde

$$\alpha_i \geqslant 0, \ \beta_i \geqslant 0 \quad (i = 1, \ \ldots, \ s),$$

$$\alpha_1 + \ldots + \alpha_s = \beta_1 + \ldots + \beta_s = 1.$$

Para un punto cualquiera $X \in [A, B]$ tenemos

$$\overline{OX} = \lambda \cdot \overline{OA} + \mu \cdot \overline{OB} \quad (\lambda \geqslant 0, \ \mu \geqslant 0 \ \text{y} \ \lambda + \mu = 1)$$

y, por consiguiente,

$$\overline{OX} = (\lambda\alpha_1 + \mu\beta_1) \cdot \overline{OM_1} + \ldots + (\lambda\alpha_s + \mu\beta_s) \cdot \overline{OM_s}.$$

Puesto que aquí

$$\lambda\alpha_i + \mu\beta_i \geqslant 0 \quad (i = 1, \ \ldots, \ s)$$

y además

$$\sum (\lambda\alpha_i + \mu\beta_i) = \lambda \sum \alpha_i + \mu \sum \beta_i = 1,$$

resulta que $X \in \mathfrak{M}_c$ y el conjunto $\mathfrak{M}_c$ es convexo.

Resta demostrar que todo conjunto convexo $\mathfrak{N}$ que contiene al conjunto $\mathfrak{M}$ contiene también todos los puntos X que satisfacen las condiciones (3), (4) y (5). En estas condiciones figura el número

natural s y para $s=1$ la afirmacion que acabamos de enunciar es trivial. Apliquemos ahora la inducción según s. Supongamos que para un valor de s cualquier punto X que satisface las condiciones (3), (4) y (5) pertenece a $\mathfrak{N}$. Consideremos el punto Y que satisfaga las condiciones

$$\overrightarrow{OY}=\alpha_1\cdot\overrightarrow{OM_1}+\ldots+\alpha_s\cdot\overrightarrow{OM_s}+\alpha_{s+1}\cdot\overrightarrow{OM_{s+1}} \quad (M_i\in\mathfrak{M}),$$
$$\alpha_1+\ldots+\alpha_s+\alpha_{s+1}=1,\ \alpha_i\geqslant 0 \quad (i=1,\ \ldots,\ s+1).$$

Siendo aquí $\alpha_s=0$ o $\alpha_{s+1}=0$, tenemos, por la hipótesis de inducción, $Y\in\mathfrak{N}$. Supongamos por esto que $\alpha_s\cdot\alpha_{s+1}\neq 0$, de modo que tomando $\mu=\alpha_1+\ldots+\alpha_s$ tendremos $\mu>0$. El punto X definido por la relación

$$\overrightarrow{OX}=\frac{\alpha_1}{\mu}\overrightarrow{OM_1}+\ldots+\frac{\alpha_s}{\mu}\overrightarrow{OM_s}$$

satisface las condiciones (3), (4) y (5) con s fijo y, por consiguiente, $X\in\mathfrak{N}$. Tenemos ahora

$$\overrightarrow{OY}=\mu\cdot\overrightarrow{OX}+\alpha_{s+1}\cdot\overrightarrow{OM_{s+1}} \quad (\mu+\alpha_{s+1}=1,\ \mu\geqslant 0\ \text{y}\ \alpha_{s+1}\geqslant 0),$$

de modo que $Y\in[X,\ M_{s+1}]$. Como $\mathfrak{N}$ es convexo y $X,\ M_{s+1}\in\mathfrak{N}$, tenemos $Y\in\mathfrak{N}$ que es lo que se quería demostrar.

En cada una de las condiciones (3) figura solamente un número finito de puntos del conjunto. Por ello, del teorema 1 resulta que *la adherencia convexa de un conjunto infinito de puntos es la unión de las adherencias convexas de todos los subconjuntos finitos del conjunto dado.*

Se dice que *la dimensión de un conjunto convexo* $\mathfrak{S}$ *es igual a* r, si $\mathfrak{S}$ está contenido en un plano r-dimensional y no está contenido en ningún plano de menor dimensión. Análogamente se define también la codimensión de un conjunto convexo.

COROLARIO. *Si el número máximo de puntos linealmente independientes del conjunto* $\mathfrak{M}$ *es igual a* $r+1$, *la dimensión de* $\operatorname{Conv}\mathfrak{M}$ *es igual a* r.

Supongamos, por ejemplo, que todos los puntos del conjunto $\mathfrak{M}$ dependen linealmente de los puntos linealmente independientes $A_0,\ A_1,\ \ldots,\ A_r$ de este conjunto. Entonces $\mathfrak{M}$ está contenido en el plano $\mathfrak{N}=A_0\vee A_1\vee\ldots\vee A_r$ de dimensión r y por ello $\operatorname{Conv}\mathfrak{M}\subseteq\operatorname{Conv}\mathfrak{N}$. Pero un plano es un conjunto convexo y, por consiguiente, $\operatorname{Conv}\mathfrak{M}\subseteq\mathfrak{N}$.

Un conjunto convexo $\mathfrak{S}$ se llama *propio* en el espacio $\mathfrak{A}$, si la codimensión de $\mathfrak{S}$ es igual a cero. Los demás conjuntos convexos se llaman *impropios* en $\mathfrak{A}$. Como ejemplos de conjuntos convexos propios podemos indicar el mismo espacio $\mathfrak{A}$, así omo los semiespacios y los semiespacios abiertos de éste. Los planos de codimensión no nula y sus semiplanos son conjuntos convexos impropios en $\mathfrak{A}$.

Está claro que todo conjunto convexo $\mathfrak{S}$ es un conjunto convexo propio en el plano que sirve de adherencia lineal de $\mathfrak{S}$.

Se dice que un punto S se halla *estrictamente dentro* del segmento $[A, B]$, si S es diferente de A y de B y $S \in [A, B]$. Un punto S se llama interior de un conjunto convexo $\mathfrak{S}$, si en toda recta que pasa por S existe un segmento que contiene a S estrictamente dentro de sí y pertenece íntegramente a $\mathfrak{S}$.

Se dice que un punto S *es tangente* a un conjunto convexo $\mathfrak{S}$, si existe una recta que pasa por S tal que cualquier segmento de ella que contenga al punto S estrictamente dentro de sí contenga al menos un punto de $\mathfrak{S}$. En particular, todos los puntos del conjunto $\mathfrak{S}$ son tangentes a $\mathfrak{S}$. Pero pueden también existir puntos tangentes a $\mathfrak{S}$ que no pertenezcan a $\mathfrak{S}$. Si todos los puntos tangentes a $\mathfrak{S}$ pertenecen a $\mathfrak{S}$, se dice que $\mathfrak{S}$ es un conjunto *cerrado*. Un conjunto convexo $\mathfrak{S}$ se llama *abierto*, si todo punto S suyo es un punto interior de $\mathfrak{S}$.

TEOREMA 2. *La intersección de una familia arbitraria de conjuntos convexos cerrados es un conjunto convexo cerrado. La intersección de una familia finita cualquiera de conjuntos convexos abiertos es un conjunto convexo abierto.*

La primera afirmación se deduce directamente de la definición de los conjuntos cerrados. Demostremos la segunda afirmación. Sean $\mathfrak{M}$ y $\mathfrak{N}$ unos conjuntos convexos abiertos. El conjunto $\mathfrak{M} \cap \mathfrak{N}$ es entonces convexo. Si es vacío, no hay nada que demostrar, ya que los conjuntos vacíos son, por definición, abiertos. Sea $S \in \mathfrak{M} \cap \mathfrak{N}$ y sea $\mathfrak{B}$ una recta cualquiera del espacio $\mathfrak{A}$ que pasa por el punto S. Sobre esta recta existen, por hipótesis, unos segmentos que contienen a S estrictamente dentro de si y que pertenecen a los respectivos conjuntos $\mathfrak{M}$ y $\mathfrak{N}$. Pero la intersección de estos segmentos será entonces el segmento deseado que pertenece a $\mathfrak{M} \cap \mathfrak{N}$ y que contiene al punto S estrictamente dentro de sí.

Introduzcamos otro concepto importante. Un punto S se llama *punto frontera* de un conjunto convexo $\mathfrak{S}$, si S es tangente a $\mathfrak{S}$, pero no es interior de $\mathfrak{S}$. En otras palabras, se dice que S es un punto frontera de $\mathfrak{S}$, si existe una recta que pasa por S tal que todo segmento de la misma que contenga al punto S estrictamente dentro de sí contenga al menos un punto de $\mathfrak{S}$ y un punto que no sea de $\mathfrak{S}$. El conjunto de todos los puntos frontera de un conjunto convexo $\mathfrak{S}$ se llama *frontera* del conjunto $\mathfrak{S}$.

TEOREMA 3. *Un semiespacio arbitrario $\mathfrak{A}_{\mathfrak{P}A}$, definido en el espacio $\mathfrak{A}$ por un hiperplano $\mathfrak{P}$ y por un punto A que no se halla sobre $\mathfrak{P}$, así como el semiespacio abierto correspondiente $\mathfrak{A}^0_{\mathfrak{P}A} = \mathfrak{A}_{\mathfrak{P}A} \setminus \mathfrak{P}$, son unos conjuntos convexos cerrado y abierto, respectivamente, en $\mathfrak{A}$. La frontera de ambos conjuntos es el hiperplano $\mathfrak{P}$. Todo conjunto convexo $\mathfrak{S}$ impropio en $\mathfrak{A}$ no contiene puntos interiores.*

Sea $P \in \mathfrak{P}$. Consideremos la recta $A \vee P$ y en ella un segmento arbitrario $[B, C]$ que contiene al punto P estrictamente dentro de sí. Puesto que la recta $A \vee P$ no se halla sobre $\mathfrak{P}$, los extremos del segmento $[B, C]$ tampoco pertenecen a $\mathfrak{P}$ y, como $P \in [B, C]$, los puntos B y C se hallan a distintos lados del hiperplano $\mathfrak{P}$. Por consiguiente, uno de los extremos B o C pertenece al semiespacio abierto $\mathfrak{A}^0_{\mathfrak{P}A}$ y el otro extremo no pertenece a $\mathfrak{A}_{\mathfrak{P}A}$. Luego, el punto P es un punto frontera tanto de $\mathfrak{A}_{\mathfrak{P}A}$ como de $\mathfrak{A}^0_{\mathfrak{P}A}$.

Demostremos que cualquier punto $D \notin \mathfrak{A}_{\mathfrak{P}A}$ no es tangente a $\mathfrak{A}_{\mathfrak{P}A}$ y, por consiguiente, tampoco es tangente a $\mathfrak{A}^0_{\mathfrak{P}A}$. Sea $\mathfrak{M}$ una recta cualquiera que pasa por el punto D. Si esta recta no se corta con $\mathfrak{P}$, pertenece íntegramente al semiespacio $\mathfrak{A}^0_{\mathfrak{P}D} = \mathfrak{A}_{\mathfrak{P}D} / \mathfrak{P}$ y, por ello, todo segmento de la misma no contiene puntos de $\mathfrak{A}_{\mathfrak{P}A}$. Si $\mathfrak{M}$ se corta con $\mathfrak{P}$ en un punto P, existen en $\mathfrak{M}$ unos puntos B y C diferentes de D y P tales que $D \in [B, P]$ y $C \in [D, P]$. Por consiguiente, el segmento $[B, C]$ se halla sobre $\mathfrak{M}$, contiene a D estrictamente dentro de sí y no contiene puntos del semiespacio $\mathfrak{A}_{\mathfrak{P}A}$. Por esto, el punto D no es tangente a $\mathfrak{A}_{\mathfrak{P}A}$. Al mismo tiempo hemos demostrado que todos los puntos del semiespacio abierto $\mathfrak{A}^0_{\mathfrak{P}A}$ son interiores y, por consiguiente, cualquier semiespacio abierto es un conjunto convexo abierto.

Para demostrar la última afirmación del teorema es suficiente indicar por $\mathfrak{P}$ el hiperplano que contiene, por hipótesis, al conjunto $\mathfrak{S}$ y observar que todo segmento de la recta $A \vee P$ ($P \in \mathfrak{S}$ y $A \notin \mathfrak{P}$) que contiene estrictamente dentro de sí al punto P contiene también necesariamente puntos que no pertenecen al hiperplano $\mathfrak{P}$.

De los teoremas 2 y 3 obtenemos un corolario importante.

COROLARIO 1. *La intersección de cualquier familia de semiespacios de un espacio $\mathfrak{A}$ es un conjunto convexo cerrado de $\mathfrak{A}$. La intersección de cualquier familia finita de semiespacios abiertos del espacio $\mathfrak{A}$ es un conjunto convexo abierto de $\mathfrak{A}$.*

Recordando que cualquiera que sea la funcional lineal no constante $f(X)$ el conjunto de soluciones de la desigualdad $f(X) \geqslant 0$ es un semiespacio y el conjunto de soluciones de la desigualdad estricta $f(X) > 0$ es un semiespacio abierto, obtenemos otro corolario.

COROLARIO 2 *Sean $f_1(X), \ldots, f_s(X)$ unas funcionales lineales no constantes sobre un espacio afín $\mathfrak{A}$. Entonces el conjunto de las soluciones X del sistema de desigualdades*

$$f_i(X) \geqslant 0 \quad (i = 1, \ldots, s)$$

es un conjunto convexo cerrado de $\mathfrak{A}$ y el conjunto de las soluciones del sistema de desigualdades estrictas

$$f_i(X) > 0 \quad (i = 1, \ldots, s)$$

es un conjunto convexo abierto de $\mathfrak{A}$.

Un hiperplano $\mathfrak{P}$ se llama hiperplano *soporte* de un conjunto convexo $\mathfrak{S}$, si $\mathfrak{P} \cap \mathfrak{S} \neq \varnothing$ y todos los puntos de $\mathfrak{S}$ se hallan a un mismo lado de $\mathfrak{P}$. Las intersecciones de los hiperplanos soportes con el conjunto $\mathfrak{S}$ se llaman *facetas* de $\mathfrak{S}$. De estas definiciones se desprende directamente que las facetas de los conjuntos convexos pertenecen a las fronteras de estos conjuntos y son ellas mismas conjuntos convexos. Las facetas de los conjuntos convexos cerrados son conjuntos cerrados.

Complementos y ejemplos

Sea $\mathfrak{A}$ un espacio afín arbitrario de dimensión finita $n \geqslant 1$ sobre un cuerpo conmutativo ordenado K.

1. La adherencia convexa de unos puntos $A_1, \ldots, A_{r+1}$ linealmente independientes se llama *simplice r-dimensional* de $\mathfrak{A}$. Si $r < n$, este simplice se halla en el plano r-dimensional $A_1 \vee \ldots \vee A_{r+1}$ y por lo tanto es un conjunto convexo *impropio* en $\mathfrak{A}$. Los simplices n-dimensionales se llaman simplemente *simplices* de $\mathfrak{A}$. Los puntos $A_1, \ldots, A_{n+1}$ se llaman vértices del simplice $\operatorname{Conv}\{A_1, \ldots, A_{n+1}\}$.

2. Describanse los simplices de los espacios unidimensionales, bidimensionales y tridimensionales.

3. Para todo r, $0 \leqslant r \leqslant n-1$, todas las facetas r-dimensionales de un simplice $A = \operatorname{Conv}\{A_1, \ldots, A_{n+1}\}$ son unos simplices $\operatorname{Conv}\{A_{i_1}, \ldots, A_{i_{r+1}}\}$ $(1 \leqslant i_1 < \ldots < i_{r+1} \leqslant n+1)$.

4. Todo conjunto convexo $\mathfrak{S}$ compuesto sólo de puntos frontera de un simplice A pertenece íntegramente a un simplice $(n-1)$-dimensional $A' = \operatorname{Conv}\{A_1, \ldots, A_{l-1}, A_{l+1}, \ldots, A_{n+1}\}$.

5. Si los puntos $A_1, \ldots, A_{r+1}$ son linealmente independientes y para unos puntos $B_1, \ldots, B_{r+1}$ se tiene

$$\operatorname{Conv}\{A_1, \ldots, A_{r+1}\} = \operatorname{Conv}\{B_1, \ldots, B_{r+1}\},$$

los conjuntos $\{A_1, \ldots, A_{r+1}\}$ y $\{B_1, \ldots, B_{r+1}\}$ coinciden.

6. La adherencia convexa $C = \operatorname{Conv}\{C_1, \ldots, C_s\}$ de un sistema finito arbitrario de puntos $C_1, \ldots, C_s$ de $\mathfrak{A}$ se llama *poliedro convexo* de $\mathfrak{A}$.

Un sistema de puntos se llama *convexamente irreducible*, si ningún punto del sistema no está contenido en la adherencia convexa de los demás puntos.

Demuéstrese que siendo $C = \operatorname{Conv}\{C_1, \ldots, C_s\}$ un poliedro convexo, existen unos puntos $C_{i_1}, \ldots, C_{i_l}$, $1 \leqslant i_1 < \ldots < i_l \leqslant s$ tales que el sistema de los puntos $C_{i_1}, \ldots, C_{i_l}$ es convexamente irreducible y

$$\operatorname{Conv}\{C_1, \ldots, C_s\} = \operatorname{Conv}\{C_{i_1}, \ldots, C_{i_l}\}$$

7. Un simplice S se llama simplice *diagonal* de un poliedro convexo s-dimensional C generado por un sistema convexamente irreducible de puntos $C_1, \ldots, C_{s+1}$, si los vértices del simplice pertenecen al conjunto $\{C_1, \ldots, C_{s+1}\}$.

Se dice que un simplice r-dimensional A y un simplice s-dimensional B están en posición regular (uno con respecto al otro), si o no se cortan o se cortan por un conjunto que es una faceta tanto de un simplice como del otro.

Tiene lugar la siguiente proposición:

Existe un conjunto de simplices diagonales de un poliedro $\mathfrak{S}$ tal que todo punto del poliedro pertenece a la unión de estos simplices y cualesquiera dos simplices de este conjunto están en posición regular.

§ 32. Espacios euclídeos puntuales

En los parágrafos anteriores hemos estudiado las propiedades de los espacios afines definidos sobre un espacio vectorial dado $\mathfrak{L}$. Supongamos ahora que el espacio vectorial $\mathfrak{L}$ es unitario o euclídeo. Un espacio afín sobre un espacio vectorial unitario (respectivamente, euclídeo) $\mathfrak{L}$ se llama espacio puntual unitario (respectivamente, euclídeo) sobre $\mathfrak{L}$. Se llama dimensión de un espacio puntual unitario sobre $\mathfrak{L}$ la dimensión del correspondiente espacio vectorial unitario $\mathfrak{L}$. En lo que sigue dedicaremos la atención principal al estudio de los espacios puntuales euclídeos sobre el cuerpo de todos los números reales. Puesto que el cuerpo de los números reales es ordenado, en los espacios puntuales euclídeos está definido el concepto de convexidad (p. 31.3).

32.1. Longitud de una quebrada. Sea $\mathfrak{U}_n$ un espacio puntual unitario de n dimensiones sobre un espacio vectorial unitario $\mathfrak{L}$ definido sobre el cuerpo conmutativo K. A todo par de puntos A y B de $\mathfrak{U}_n$ corresponde un vector univoco $\overline{AB}$ de $\mathfrak{L}$. Puesto que el espacio vectorial $\mathfrak{L}$ es unitario, en $\mathfrak{L}$ está definido el concepto de longitud del vector $\overline{AB}$. Indicaremos por $\rho(A, B)$ esta longitud llamándola *distancia* del punto A al punto B. Por consiguiente, tenemos por definición

$$\rho(A, B) = \|\overline{AB}\| = \sqrt{(\overline{AB}, \overline{AB})}. \qquad (1)$$

De las propiedades de las longitudes de los vectores (p. 17.2) se desprende que para cualesquiera puntos A y B se tiene

1° $\qquad \rho(A, B) = \rho(B, A).$

2° $\qquad \rho(A, B) = 0 \Leftrightarrow A = B.$

Como para cualesquiera puntos A, B y C de $\mathfrak{U}_n$ se tiene $\overline{AC} = \overline{AB} + \overline{BC}$, de la estimación para la longitud de una suma de vectores resulta

3° $\qquad \rho(A, B) + \rho(B, C) \geqslant \rho(A, C).$

Un conjunto arbitrario M tal que a todo par A, B de sus elementos corresponde un número real no negativo $\rho(A, B)$ que satisface las exigencias 1°, 2° y 3° se llama *espacio métrico* de métrica ρ. Por esto la definición (1) convierte un espacio puntual unitario en un espacio métrico.

Aplicando sucesivamente la desigualdad 3°, llegamos fácilmente a la desigualdad más general

$$\rho(A_1, A_2) + \rho(A_2, A_3) + \ldots + \rho(A_{s-1}, A_s) \geqslant \rho(A_1, A_s), \qquad (2)$$

válida para cualquier sucesión finita de puntos $A_1, \ldots, A_s$ de un espacio puntual unitario.

Si un espacio puntual Π_n es euclídeo, el campo principal K — que es el cuerpo de los números reales — es ordenado y, por ello, todo par de puntos A y B de Π_n determina un segmento $[A, B]$ (p. 31.1). El número $\rho(A, B)$ se llama *longitud* del segmento $[A, B]$. Dada una sucesión cualquiera de puntos $A_1, A_2, \ldots, A_s$, la sucesión de los segmentos

$$[A_1, A_2], [A_2, A_3], \ldots, [A_{s-1}, A_s]$$

se llama *quebrada* que une A_1 y A_s y el segmento $[A_1, A_s]$ se llama segmento *resultante* de la quebrada. La suma de las longitudes de los segmentos de una quebrada se llama *longitud de la quebrada*. La desigualdad (2) significa que en un espacio euclídeo cualquiera la longitud de una quebrada es no menor que la longitud de su segmento resultante.

¿Bajo qué condiciones en la desigualdad (2) tiene lugar el signo de igualdad? Según el p. 17.2 (para $s=2$), en un espacio vectorial unitario de las relaciones $\overline{A_1A_2} \neq o$ y

$$\| \overline{A_1A_2} \| + \| \overline{A_2A_3} \| + \ldots + \| \overline{A_{s-1}A_s} \| = \| \overline{A_1A_2} + \ldots + \overline{A_{s-1}A_s} \| \quad (3)$$

se deduce que los vectores A_iA_{i+1} dependen linealmente de $\overline{A_1A_2}$ y por lo tanto

$$\overline{A_iA_{i+1}} = \lambda_i \cdot \overline{A_1A_2} \qquad (i = 1, \ldots, s-1).$$

Introduciendo estos valores en (3) y dividiendo por $\| \overline{A_1A_2} \|$, llegamos a la igualdad

$$|\lambda_1| + |\lambda_2| + \ldots + |\lambda_{s-1}| = |\lambda_1 + \lambda_2 + \ldots + \lambda_{s-1}|. \quad (4)$$

Si el espacio es euclídeo, los números $\lambda_1, \ldots, \lambda_{s-1}$ son reales y la igualdad (4) es verídica cuando, y sólo cuando, estos números son de un mismo signo. Está claro que en este caso los puntos $A_1, A_2, \ldots, A_s$ se hallan sobre una misma recta y que los segmentos sucesivos $[A_1, A_2], [A_2, A_3], \ldots, [A_{s-1}, A_s]$ de la quebrada se intersecan sólo en los correspondientes puntos extremos. Es decir, *en un espacio puntual euclídeo la longitud de una quebrada es igual a la longitud del segmento resultante cuando, y sólo cuando, la quebrada es una partición del segmento resultante.*

Veamos ahora como determinar la distancia entre los puntos A y B de un espacio puntual unitario Π_n si se conocen las coordenadas de estos puntos. Sea $(O, e_1, \ldots, e_n)$ una base de referencia coordenada de Π_n cuyos vectores forman un sistema ortonormal en el espacio vectorial $\mathfrak{L}$. Indicando por $\alpha_1, \ldots, \alpha_n$ y $\beta_1, \ldots, \beta_n$ las coordenadas de los puntos A y B en la base de referencia señalada, tendremos

$$\overrightarrow{AB} = \overrightarrow{OB} - \overrightarrow{OA} = (\beta_1 - \alpha_1) e_1 + \ldots + (\beta_n - \alpha_n) e_n$$

de modo que

$$\rho(A, B) = \sqrt{|\beta_1 - \alpha_1|^2 + \ldots + |\beta_n - \alpha_n|^2}. \quad (5)$$

Si el espacio $\mathfrak{U}_n$ es euclídeo, las coordenadas de los puntos serán números reales y, por consiguiente, la fórmula (5) puede ser representada en la forma

$$\rho(A, B) = \sqrt{(\beta_1 - \alpha_1)^2 + \ldots + (\beta_n - \alpha_n)^2}. \quad (6)$$

Al deducir estas fórmulas hemos aceptado que los vectores de la base de referencia coordenada forman un sistema ortonormal. Estas bases de referencia se llamarán *ortonormales*. Si en lugar de una base de referencia coordenada ortonormal se toma una cualquiera, la fórmula para la distancia entre dos puntos adquiere una forma más compleja que aquí no la daremos.

32.2. Angulo entre rectas. Consideremos en un espacio puntual unitario $\mathfrak{U}_n$ de n dimensiones dos rectas cualesquiera $\mathfrak{X}$ y $\mathfrak{Y}$. Tomemos en cada una de estas rectas un par de diferentes puntos A_1, A_2 y A_3, A_4, respectivamente. Los vectores $\overline{A_1A_2}$ y $\overline{A_3A_4}$ que corresponden a estos pares satisfacen la desigualdad de Cauchy—Buniakovski

$$\frac{|(\overline{A_1A_2}, \overline{A_3A_4})|}{\|\overline{A_1A_2}\| \cdot \|\overline{A_3A_4}\|} \leqslant 1.$$

Por lo tanto, existe un número real φ que satisface la exigencia

$$\cos\varphi = \frac{|(\overline{A_1A_2}, \overline{A_3A_4})|}{\|\overline{A_1A_2}\| \cdot \|\overline{A_3A_4}\|} \quad \left(0 \leqslant \varphi \leqslant \frac{\pi}{2}\right). \quad (1)$$

Este número φ no depende de cómo se escogan los pares indicados de puntos en las rectas $\mathfrak{X}$ y $\mathfrak{Y}$. Efectivamente, si B_1, B_2 y B_3, B_4 son otros pares análogos, se tiene

$$\overline{B_1B_2} = \lambda \cdot \overline{A_1A_2} \quad \text{y} \quad \overline{B_3B_4} = \mu \cdot \overline{A_3A_4}$$

para unos valores adecuados λ, $\mu \in K$ y por ello

$$\frac{|(\overline{B_1B_2}, \overline{B_3B_4})|}{\|\overline{B_1B_2}\| \cdot \|\overline{B_3B_4}\|} = \frac{|\lambda\mu|}{|\lambda| \cdot |\mu|} \frac{|(\overline{A_1A_2}, \overline{A_3A_4})|}{\|\overline{A_1A_2}\| \cdot \|\overline{A_3A_4}\|} = \cos\varphi.$$

Luego, el número real φ que cumple las exigencias (1) depende sólo de las rectas $\mathfrak{X}$ y $\mathfrak{Y}$ y se llama *magnitud del ángulo* entre las rectas $\mathfrak{X}$ y $\mathfrak{Y}$ o simplemente *ángulo* entre las rectas $\mathfrak{X}$ y $\mathfrak{Y}$.

Las rectas $\mathfrak{X}$ y $\mathfrak{Y}$ se llaman *perpendiculares*, si el ángulo entre ellas es igual a $\pi/2$. De las relaciones (1) se deduce que *las rectas $\mathfrak{X}$ y $\mathfrak{Y}$ son perpendiculares cuando, y sólo cuando, cualquier vector que se halla sobre una de estas rectas es ortogonal a cualquier vector que se halla sobre la otra recta.*

Notemos también que *el ángulo entre las rectas* $\mathfrak{X}$ *y* $\mathfrak{Y}$ *es igual a cero cuando, y sólo cuando, las rectas son paralelas.*

En efecto, si las rectas son paralelas, los vectores que se hallan sobre estas rectas son linealmente dependientes y por ello en la fórmula (1) tenemos $\overrightarrow{A_3A_4} = \lambda \cdot \overrightarrow{A_1A_2}$ y mediante simplificaciones directas obtenemos $\cos\varphi = 1$ y $\varphi = 0$. Recíprocamente, sea $\varphi = 0$ y, por consiguiente,

$$|(\overrightarrow{A_1A_2}, \overrightarrow{A_3A_4})| = \|\overrightarrow{A_1A_2}\| \cdot \|\overrightarrow{A_3A_4}\|.$$

Este es el caso en el que en la desigualdad de Cauchy—Buniakovski tiene lugar el signo de igualdad. Por consiguiente, $\overrightarrow{A_3A_4} = \lambda \cdot \overrightarrow{A_1A_2}$, de modo que las rectas $\mathfrak{X}$ y $\mathfrak{Y}$ son paralelas.

Fijemos en el espacio unitario U_n una base de referencia coordenada ortonormal $(O, e_1, \ldots, e_n)$. Como sabemos las ecuaciones de las rectas $\mathfrak{X}$ y $\mathfrak{Y}$ pueden ser representadas en la forma

$$\frac{\xi_1 - \xi_1^0}{l_1} = \ldots = \frac{\xi_n - \xi_n^0}{l_n} = t, \tag{2}$$

$$\frac{\xi_1 - \eta_1^0}{m_1} = \ldots = \frac{\xi_n - \eta_n^0}{m_n} = t, \tag{3}$$

donde t es un parámetro y $l_1, \ldots, l_n$ y $m_1, \ldots, m_n$ son los coeficientes directores. Supongamos que el punto A_i se obtiene para el valor $t = t_i$ $(i = 1, 2, 3, 4)$. Indicando por $\alpha_{i1}, \ldots, \alpha_{in}$ las coordenadas del punto $A_i (i = 1, 2, 3, 4)$, tendremos

$$\alpha_{is} = l_s t_i + \xi_s^0 \quad (i = 1, 2;\ s = 1, \ldots, n),$$
$$\alpha_{is} = m_s t_i + \eta_s^0 \quad (i = 3, 4;\ s = 1, \ldots, n)$$

y, por consiguiente,

$$\overrightarrow{A_1A_2} = (t_2 - t_1)(l_1e_1 + \ldots + l_ne_n),$$
$$\overrightarrow{A_3A_4} = (t_4 - t_3)(m_1e_1 + \ldots + m_ne_n).$$

Introduciendo en la fórmula (1) estos valores, obtenemos

$$\cos\varphi = \frac{|l_1m_1 + \ldots + l_nm_n|}{\sqrt{|l_1|^2 + \ldots + |l_n|^2}\sqrt{|m_1|^2 + \ldots + |m_n|^2}}. \tag{4}$$

Esta es la fórmula estandard para el coseno del ángulo entre unas rectas dadas por las ecuaciones canónicas (2) y (3) en un sistema de coordenadas ortonormal. Tomando la recta (3) como la recta coordenada s-ésima, definida por las ecuaciones

$$\frac{\xi_1}{0} = \ldots = \frac{\xi_s}{1} = \ldots = \frac{\xi_n}{0}$$

e indicando por φ_s el ángulo entre la recta (1) y esta recta coordenada, obtenemos

$$\cos\varphi_s = \frac{|l_s|}{\sqrt{|l_1|^2 + \ldots + |l_n|^2}}$$

y, en particular,

$$\cos^2 \varphi_1 + \ldots + \cos^2 \varphi_n = 1. \qquad (5)$$

Hemos visto ya en el p. 30.3 que la condición de paralelismo de las rectas (2) y (3) puede ser representada en la forma

$$\frac{l_1}{m_1} = \ldots = \frac{l_n}{m_n}.$$

Tomando en la fórmula (4) $\varphi = \frac{\pi}{2}$, llegamos a la relación

$$l_1 m_1 + \ldots + l_n m_n = 0, \qquad (6)$$

que representa *la condición de perpendicularidad de las rectas dadas por las ecuaciones canónicas en una base de referencia coordenada ortonormal.*

En los espacios puntuales euclídeos se introduce también, además del concepto de ángulo entre rectas, el concepto de ángulo entre rayos. Las ecuaciones canónicas de los rayos en un espacio euclídeo n-dimensional puntual también se representan en la forma (2) y (3), donde $\xi_1^0, \ldots, \xi_n^0$ y $\eta_1^0, \ldots, \eta_n^0$ son las coordenadas de los vértices de estos rayos, y las coordenadas $\xi_1, \ldots, \xi_n$ de un punto arbitrario de estos rayos se obtienen de las ecuaciones indicadas dando al parámetro t unos valores no negativos arbitrarios. Puesto que el campo principal es en este caso el cuerpo de los números reales, existe un número real único φ que cumple las exigencias

$$\cos \varphi = \frac{l_1 m_1 + \ldots + l_n m_n}{\sqrt{l_1^2 + \ldots + l_n^2}\sqrt{m_1^2 + \ldots + m_n^2}} \qquad (0 \leqslant \varphi \leqslant \pi), \quad (7)$$

que se denomina *ángulo entre los rayos* señalados. Según estas definiciones el ángulo entre cualesquiera rectas está comprendido siempre en los límites 0 y $\frac{\pi}{2}$, mientras que el ángulo entre unos rayos puede ser también obtuso. Supongamos, por ejemplo, que un rayo está dado por sus ecuaciones canónicas

$$\frac{\xi_1 - \xi_1^0}{l_1} = \ldots = \frac{\xi_n - \xi_n^0}{l_n} = t \qquad (t \geqslant 0).$$

La ecuación del rayo opuesto que con el rayo dado forma una recta integra es entonces de la forma

$$\frac{\xi_1 - \xi_1'}{l_1} = \ldots = \frac{\xi_n - \xi_n^0}{l_n} = t \qquad (t \leqslant 0)$$

o en forma canónica

$$\frac{\xi_1 - \xi_1^0}{-l_1} = \ldots = \frac{\xi_n - \xi_n^0}{-l_n} = t \qquad (t \geqslant 0)$$

Calculando el ángulo entre estos rayos opuestos según la fórmula (7), obtenemos $\cos\varphi = -1$ y por lo tanto $\varphi = \pi$.

32.3. Proyecciones ortogonales. Empleando el concepto de perpendicularidad de rectas es fácil definir la relación de perpendicularidad de unos planos k-dimensional y l-dimensional en un espacio unitario $\mathfrak{U}_n$. Se dice que la recta $\mathfrak{X}$ del espacio $\mathfrak{U}_n$ es perpendicular a un plano $\mathfrak{M}$ k-dimensional de $\mathfrak{U}_n$, lo que se indica simbólicamente por $\mathfrak{X} \perp \mathfrak{M}$, si $\mathfrak{X}$ es perpendicular a cualquier recta que pertenezca a $\mathfrak{M}$.

Puesto que la perpendicularidad de las rectas equivale a la ortogonalidad de los vectores no nulos que se hallan sobre ellas, se tiene

$$\mathfrak{X} \perp \mathfrak{M} \Leftrightarrow \overline{\mathfrak{X}} \perp \overline{\mathfrak{M}}, \tag{1}$$

donde $\overline{\mathfrak{M}}$ significa el subespacio tangente al plano $\mathfrak{M}$ (en las notaciones del p. 29.3 se tiene $\overline{\mathfrak{M}} = L(\mathfrak{M})$).

Una recta $\mathfrak{X}$ se llama perpendicular trazada desde el punto A hacia el plano $\mathfrak{M}$, si $\mathfrak{X}$ pasa por A, es perpendicular a $\mathfrak{M}$ y se corta con $\mathfrak{M}$ en un punto P. El punto P se llama base de la perpendicular trazada desde A hacia $\mathfrak{M}$ o proyección de A sobre $\mathfrak{M}$.

TEOREMA 1. *En un espacio unitario $\mathfrak{U}_n$ de n dimensiones desde todo punto A que no se halle sobre un plano arbitrario k-dimensional $\mathfrak{M}$ $(1 \leqslant k \leqslant n-1)$ se puede trazar una perpendicular hacia $\mathfrak{M}$, y sólo una.*

UNICIDAD. Supongamos que existen dos rectas diferentes $A \vee B$ y $A \vee C$ $(B \neq C;\ B,\ C \in \mathfrak{M})$ perpendiculares a $\mathfrak{M}$. Estas rectas deben ser entonces perpendiculares a la recta $B \vee C$ que se halla sobre el plano $\mathfrak{M}$.

Supongamos que las rectas $A \vee B$, $A \vee C$ y $B \vee C$—que se hallan obviamente sobre el plano bidimensional $A \vee B \vee C$—tienen, en una base de referencia ortonormal del plano $A \vee B \vee C$, los coeficientes directores (l_1, l_2), (m_1, m_2) y (n_1, n_2), respectivamente. La condición de ortogonalidad de las rectas $A \vee B$ y $B \vee C$ implica la igualdad $l_1 n_1 + l_2 n_2 = 0$ (véase (6)); análogamente la ortogonalidad de $A \vee C$ y de $B \vee C$ implica la igualdad $m_1 n_1 + m_2 n_2 = 0$. Vemos que el sistema

$$\left.\begin{aligned} l_1 x_1 + l_2 x_2 &= 0, \\ m_1 x_1 + m_2 x_2 &= 0 \end{aligned}\right\}$$

tiene una solución no trivial (n_1, n_2). Por consiguiente,

$$\begin{vmatrix} l_1 & l_2 \\ m_1 & m_2 \end{vmatrix} = 0,$$

es decir, $\frac{l_1}{m_1} = \frac{l_2}{m_2}$. Luego, las rectas $A \vee B$ y $A \vee C$ son paralelas (véase el p. 30.3) y tienen un punto común A. De aquí se des-

prende que ellas coinciden. La contradicción obtenida concluye la demostración de la unicidad.

EXISTENCIA. Supongamos primero que $\mathfrak{M}$ es un hiperplano, es decir, que $k=n-1$. En $\mathfrak{U}_n$ existe, según el p. 17.5, un subespacio vectorial de una dimensión $\overline{\mathfrak{M}}^{\perp}$ ortogonal a $\overline{\mathfrak{M}}$. El conjunto $A\cdot\overline{\mathfrak{M}}^{\perp}$ es una recta que pasa por A y es perpendicular a $\mathfrak{M}$. Resta probar solamente que $A\cdot\overline{\mathfrak{M}}^{\perp}$ se interseca con $\mathfrak{M}$. Pero, esto es evidente, ya que en el caso contrario la recta $A\cdot\overline{\mathfrak{M}}^{\perp}$ sería paralela a $\mathfrak{M}$ y por lo tanto en $\mathfrak{M}$ existiría una recta paralela (y no perpendicular) a $A\cdot\overline{\mathfrak{M}}^{\perp}$.

Si ahora $\mathfrak{M}$ no es un hiperplano de $\mathfrak{U}_n$, consideremos el subespacio $\mathfrak{N}=A\vee\mathfrak{M}$. El conjunto $\mathfrak{M}$ es un hiperplano del espacio unitario $\mathfrak{N}$. Luego, según lo anterior, desde A se puede trazar dentro de $\mathfrak{N}$ una perpendicular hacia $\mathfrak{M}$. Es obvio que esta perpendicular será también una perpendicular dentro del espacio principal $\mathfrak{U}_n$.

En el teorema (1) se resuelve el problema sobre la posibilidad de trazar desde el punto A una perpendicular de modo que *corte* el plano $\mathfrak{M}$. Veamos cuál es la situación, si omitimos esta exigencia. Según el p. 17.5, el conjunto $\overline{\mathfrak{M}}^{\perp}$ de todos los vectores del espacio $\mathfrak{U}_n$ ortogonales a $\overline{\mathfrak{M}}$ es un subespacio vectorial de dimensión $n-k$. Por esto el conjunto $A\cdot\overline{\mathfrak{M}}^{\perp}$ será el conjunto de los puntos que se hallan sobre todas las rectas que pasan por A y que son perpendiculares a $\mathfrak{M}$. Puesto que $\overline{\mathfrak{M}}^{\perp}$ es un subespacio vectorial de dimensión $n-k$, el conjunto $A\cdot\overline{\mathfrak{M}}^{\perp}$ es un plano $(n-k)$-dimensional. La intersección de $A\cdot\overline{\mathfrak{M}}^{\perp}$ y de $\mathfrak{M}$ no puede contener ninguna recta $\mathfrak{X}$, ya que en el caso contrario cualquier vector que se halla sobre $\mathfrak{X}$ pertenecería simultáneamente a los espacios $\overline{\mathfrak{M}}$ y $\overline{\mathfrak{M}}^{\perp}$, lo cual es imposible. Por consiguiente, la intersección de $\mathfrak{M}$ y $A\cdot\overline{\mathfrak{M}}^{\perp}$ o es vacía o consta de un solo punto. Si $A\in\mathfrak{M}$, la intersección de $\mathfrak{M}$ y $A\cdot\overline{\mathfrak{M}}^{\perp}$ consta, obviamente, del punto A. En cambio, si $A\notin\mathfrak{M}$, la base de la perpendicular trazada desde A hacia $\mathfrak{M}$ pertenece simultáneamente a $A\cdot\overline{\mathfrak{M}}^{\perp}$ y a $\mathfrak{M}$ y, por ello, el conjunto $\mathfrak{M}\cap A\cdot\overline{\mathfrak{M}}^{\perp}$ está formado solamente por la base de la perpendicular mencionada. El plano $A\cdot\overline{\mathfrak{M}}^{\perp}$ suele llamarse a veces *complemento ortogonal a* $\mathfrak{M}$ *trazado por el punto* A. En particular, si $\mathfrak{M}$ es un hiperplano y $A\notin\mathfrak{M}$, el complemento ortogonal $A\cdot\overline{\mathfrak{M}}^{\perp}$ es la recta que es perpendicular al hiperplano $\mathfrak{M}$ y que lo intercepta en el punto A. Esta recta se llama perpendicular al hiperplano $\mathfrak{M}$ levantada desde el punto A.

Se llama *proyección* (con más precisión, *proyección ortogonal*) de un punto A sobre un plano cualquiera $\mathfrak{M}$ el punto de intersección

de los planos $\mathfrak{M}$ y $A \cdot \overline{\mathfrak{M}}^{\perp}$. En otras palabras, si $A \notin \mathfrak{M}$, se llama proyección de A sobre $\mathfrak{M}$ la base de la perpendicular trazada desde A hacia $\mathfrak{M}$. Si $A \in \mathfrak{M}$, se llama proyección de A sobre $\mathfrak{M}$ el propio punto A. La proyección de A sobre $\mathfrak{M}$ se indica a veces por $\mathrm{Pr}_{\mathfrak{M}} A$ o por $A_{\mathfrak{M}}$. La aplicación $\mathrm{Pr}_{\mathfrak{M}}: \mathfrak{U}_n \to \mathfrak{M}$ se llama proyección (ortogonal) del espacio $\mathfrak{U}_n$ sobre el plano $\mathfrak{M}$.

Veamos cómo se expresan las coordenadas de la proyección del punto A sobre el plano $\mathfrak{M}$ en términos de las coordenadas de A. Tomemos en el espacio $\mathfrak{M}$ una base ortonormal $e_1, \ldots, e_k$ y complementémosla con los vectores $e_{k+1}, \ldots, e_n$ hasta obtener una base otronormal $e_1, \ldots, e_k, \ldots, e_n$ de todo el espacio $\mathfrak{U}_n$. Fijemos en $\mathfrak{M}$ un punto cualquiera O y tomemos como base de referencia coordenada de $\mathfrak{U}_n$ la base de referencia $(O, e_1, \ldots, e_n)$. Sean $\alpha_1, \ldots, \alpha_n$ las coordenadas del punto A en esta base de referencia. Demostremos que la proyección de A sobre $\mathfrak{M}$ será el punto P con las coordenadas $\alpha_1, \ldots, \alpha_k, 0, \ldots, 0$ (véase la fig. 10 para $k=2$). En efecto, el plano $\mathfrak{M}$ consta de los puntos cuyas filas coordenadas son de la forma $(\xi_1, \ldots, \xi_k, 0, \ldots, 0)$. Por esto todo vector x perteneciente a $\overline{\mathfrak{M}}$ puede ser representado en la forma $\xi_1 e_1 + \ldots + \xi_k e_k$. Pero $\overline{PA} = \alpha_{k+1} e_{k+1} + \ldots + \alpha_n e_n$ y por lo tanto se tiene $(x, \overline{PA}) = 0$, es decir, la recta $P \vee A$ es perpendicular al plano $\mathfrak{M}$ que es lo que se quería demostrar. Luego, si

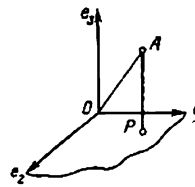


Fig. 10.

$$\overline{OA} = \alpha_1 e_1 + \ldots + \alpha_k e_k + \ldots + \alpha_n e_n,$$

resulta

$$\overline{OA}_{\mathfrak{M}} = \alpha_1 e_1 + \ldots + \alpha_k e_k, \tag{2}$$

donde

$$\mathfrak{M} = O \vee Oe_1 \vee \ldots \vee Oe_k.$$

De la fórmula (2) se desprende directamente el siguiente teorema importante.

TEOREMA 2. *La proyección ortogonal* $\mathrm{Pr}_{\mathfrak{M}}: \mathfrak{U}_n \to \mathfrak{M}$ *es una aplicación lineal, es decir, para cualesquiera puntos* $A^{(1)}, A^{(2)}, A^{(3)}, A^{(4)}, A^{(5)}$ *y* $A^{(6)}$ *y para cualesquiera números* $\lambda, \mu \in K$ *la relación*

$$\overline{A^{(1)}A^{(2)}} = \lambda \cdot \overline{A^{(3)}A^{(4)}} + \mu \cdot \overline{A^{(5)}A^{(6)}} \tag{3}$$

implica

$$\overline{A^{(1)}_{\mathfrak{M}} A^{(2)}_{\mathfrak{M}}} = \lambda \cdot \overline{A^{(3)}_{\mathfrak{M}} A^{(4)}_{\mathfrak{M}}} + \mu \cdot \overline{A^{(5)}_{\mathfrak{M}} A^{(6)}_{\mathfrak{M}}}. \tag{4}$$

Efectivamente, indicando por $\alpha_1^{(i)}, \ldots, \alpha_n^{(i)}$ las coordenadas del punto $A^{(i)}$ $(i = 1, \ldots, 6)$ y tomando en consideración la condición

(3), obtenemos las igualdades

$$\alpha_s^{(2)} - \alpha_s^{(1)} = \lambda(\alpha_s^{(4)} - \alpha_s^{(3)}) + \mu(\alpha_s^{(6)} - \alpha_s^{(5)}) \quad (s = 1, \ldots, n). \quad (5)$$

Por otra parte, debido a (2), tenemos

$$\overline{A_{\mathfrak{M}}^{(i)} A_{\mathfrak{M}}^{(i+1)}} = (\alpha_1^{(i+1)} - \alpha_1^{(i)})\, e_1 + \ldots + (\alpha_k^{(i+1)} - \alpha_k^{(i)})\, e_k \quad (i = 2, 4, 6). \quad (6)$$

Comparando (6) y (5) llegamos a la relación (4).

Se llama proyección $\mathfrak{S}_{\mathfrak{M}}$ de un conjunto arbitrario de puntos $\mathfrak{S}$ sobre un plano $\mathfrak{M}$ el conjunto de las proyecciones sobre $\mathfrak{M}$ de todos los puntos de $\mathfrak{S}$.

TEOREMA 3. *La proyección de una recta $\mathfrak{X}$ sobre un plano cualquiera $\mathfrak{M}$ es una recta o un punto. Si $\mathfrak{X}^{(1)}, \ldots, \mathfrak{X}^{(s)}$ son unas rectas de un espacio unitario $\mathfrak{U}_n$ y $\mathfrak{M}$ es un plano cualquiera de $\mathfrak{U}_n$, se tiene*

$$(\mathfrak{X}^{(1)} \vee \ldots \vee \mathfrak{X}^{(s)})_{\mathfrak{M}} = \mathfrak{X}_{\mathfrak{M}}^{(1)} \vee \ldots \vee \mathfrak{X}_{\mathfrak{M}}^{(s)}. \quad (7)$$

Tomemos en cada una de las rectas $\mathfrak{X}^{(i)}$ un par de puntos diferentes $A^{(i)}$, $B^{(i)}$ $(i = 1, \ldots, s)$. El plano $\mathfrak{X}^{(1)} \vee \ldots \vee \mathfrak{X}^{(s)}$ está formado por aquellos puntos C para los cuales

$$\overline{A^{(1)}C} = \lambda_1 \cdot \overline{A^{(1)}B^{(1)}} + \ldots + \lambda_s \cdot \overline{A^{(1)}B^{(s)}} + \\ + \mu_2 \cdot \overline{A^{(1)}A^{(2)}} + \ldots + \mu_s \cdot \overline{A^{(1)}A^{(s)}} \quad (8)$$

para unos valores adecuados $\lambda_1, \ldots, \lambda_s, \mu_2, \ldots, \mu_s \in K$ (véase el p. 29.2). Basándonos en (4), obtenemos de aquí

$$\overline{A_{\mathfrak{M}}^{(1)}C_{\mathfrak{M}}} = \lambda_1 \cdot \overline{A_{\mathfrak{M}}^{(1)}B_{\mathfrak{M}}^{(1)}} + \ldots + \lambda_s \cdot \overline{A_{\mathfrak{M}}^{(1)}B_{\mathfrak{M}}^{(s)}} + \\ + \mu_2 \cdot \overline{A_{\mathfrak{M}}^{(1)}A_{\mathfrak{M}}^{(2)}} + \ldots + \mu_s \cdot \overline{A_{\mathfrak{M}}^{(1)}A_{\mathfrak{M}}^{(s)}}, \quad (9)$$

de modo que

$$C_{\mathfrak{M}} = (A_{\mathfrak{M}}^{(1)} \vee B_{\mathfrak{M}}^{(1)}) \vee \ldots \vee (A_{\mathfrak{M}}^{(s)} \vee B_{\mathfrak{M}}^{(s)}).$$

Recíprocamente, si

$$D \in (A_{\mathfrak{M}}^{(1)} \vee B_{\mathfrak{M}}^{(1)}) \vee \ldots \vee (A_{\mathfrak{M}}^{(s)} \vee B_{\mathfrak{M}}^{(s)}),$$

para valores adecuados $\lambda_1, \ldots, \lambda_s, \mu_2, \ldots, \mu_s \in K$ se tiene

$$\overline{A_{\mathfrak{M}}^{(1)}D} = \lambda_1 \cdot \overline{A_{\mathfrak{M}}^{(1)}B_{\mathfrak{M}}^{(1)}} + \ldots + \lambda_s \cdot \overline{A_{\mathfrak{M}}^{(1)}B_{\mathfrak{M}}^{(s)}} + \\ + \mu_2 \cdot \overline{A_{\mathfrak{M}}^{(1)}A_{\mathfrak{M}}^{(2)}} + \ldots + \mu_s \cdot \overline{A_{\mathfrak{M}}^{(1)}A_{\mathfrak{M}}^{(s)}}. \quad (10)$$

Tomemos en el espacio $\mathfrak{X}^{(1)} \vee \ldots \vee \mathfrak{X}^{(s)}$ un punto C tal que se cumpla la igualdad (8). Comparando (9) y (10) vemos que $C_{\mathfrak{M}} = D$ y por ello $D \in (\mathfrak{X}^{(1)} \vee \ldots \vee \mathfrak{X}^{(s)})_{\mathfrak{M}}$. Vemos, por consiguiente, que

$$(\mathfrak{X}^{(1)} \vee \ldots \vee \mathfrak{X}^{(s)})_{\mathfrak{M}} = (A_{\mathfrak{M}}^{(1)} \vee B_{\mathfrak{M}}^{(1)}) \vee \ldots \vee (A_{\mathfrak{M}}^{(s)} \vee B_{\mathfrak{M}}^{(s)}).$$

Para $s = 1$ obtenemos $\mathfrak{X}_{\mathfrak{M}}^{(1)} = A_{\mathfrak{M}}^{(1)} \vee B_{\mathfrak{M}}^{(1)}$. Por consiguiente $\mathfrak{X}_{\mathfrak{M}}^{(i)} = A_{\mathfrak{M}}^{(i)} \vee B_{\mathfrak{M}}^{(i)}$ y quedan demostradas ambas afirmaciones del teorema 3.

COROLARIO. *En un espacio unitario la proyección de un plano* $\mathfrak{N}$ *sobre otro plano es de nuevo un plano cuya dimensión no sobrepasa la dimensión de* $\mathfrak{N}$.

Para la demostración es suficiente representar el plano $\mathfrak{N}$ en forma de la adherencia lineal de las rectas independientes que pasan por un punto y aplicar la fórmula (7).

En los espacios puntuales euclídeos están definidos los conceptos de segmento y de cuerpo convexo. Mediante la fórmula (4) se comprueba fácilmente el teorema siguiente.

TEOREMA 4. *En un espacio puntual euclídeo la proyección de un segmento cualquiera* $[A, B]$ *sobre un plano* $\mathfrak{N}$ *es un segmento* $[A_{\mathfrak{N}}, B_{\mathfrak{N}}]$ *y por ello la proyección de cualquier conjunto convexo es un conjunto convexo.*

En efecto, si $C \in [A, B]$, se tiene $\overline{AC} = \lambda \cdot \overline{AB}$, $0 \leqslant \lambda \leqslant 1$. De aquí se deduce, debido a (4), que $\overline{A_{\mathfrak{N}}C_{\mathfrak{N}}} = \lambda \cdot \overline{A_{\mathfrak{N}}B_{\mathfrak{N}}}$ y por lo tanto $C_{\mathfrak{N}} \in [A_{\mathfrak{N}}, B_{\mathfrak{N}}]$ y $[A, B]_{\mathfrak{N}} \subseteq [A_{\mathfrak{N}}, B_{\mathfrak{N}}]$. Recíprocamente, si $D \in [A_{\mathfrak{N}}, B_{\mathfrak{N}}]$, para un valor conveniente de λ $(0 \leqslant \lambda \leqslant 1)$ tenemos $\overline{A_{\mathfrak{N}}D} = \lambda \cdot \overline{A_{\mathfrak{N}}B_{\mathfrak{N}}}$. Tomando en el segmento $[A, B]$ un punto C tal que $\overline{AC} = \lambda \cdot \overline{AB}$ y comparando (4) con la igualdad $\overline{A_{\mathfrak{N}}D} = \lambda \cdot \overline{A_{\mathfrak{N}}B_{\mathfrak{N}}}$, obtenemos $D = C_{\mathfrak{N}}$ y, por consiguiente, $[A_{\mathfrak{N}}, B_{\mathfrak{N}}] \subseteq [A, B]_{\mathfrak{N}}$ que es lo que se quería demostrar.

Veamos cómo pueden calcularse las coordenadas de la proyección de un punto dado sobre un hiperplano dado, si se conocen las coordenadas del punto y la ecuación del hiperplano. Sea $(O, e_1, \ldots, e_n)$ una base de referencia coordenada ortonormal de un espacio unitario, sea B un punto de coordenadas $\beta_1, \ldots, \beta_n$ y sea

$$\alpha_1 \xi_1 + \ldots + \alpha_n \xi_n = \alpha \tag{11}$$

la ecuación del hiperplano dado. Consideremos dos puntos distintos cualesquiera C y D de coordenadas respectivas $\gamma_1, \ldots, \gamma_n$ y $\delta_1, \ldots, \delta_n$ que se hallan sobre el hiperplano (11). Tenemos entonces

$$\overline{CD} = (\delta_1 - \gamma_1) e_1 + \ldots + (\delta_n - \gamma_n) e_n. \tag{12}$$

Las coordenadas de C y de D deben satisfacer la ecuación (11). Introduciéndolas en la ecuación (11) y restando término por término las dos igualdades obtenidas, encontramos

$$\alpha_1 (\delta_1 - \gamma_1) + \ldots + \alpha_n (\delta_n - \gamma_n) = 0. \tag{13}$$

De (12) y (13) vemos que la recta

$$\frac{\xi_1 - \beta_1}{\alpha_1} = \ldots = \frac{\xi_n - \beta_n}{\alpha_n} \tag{14}$$

es perpendicular a cualquier vector $\overline{CD}$ que se halle sobre el plano

(11) y, por ende, es perpendicular a este plano. Puesto que la recta (14) pasa por el punto B, *las igualdades* (14) *son las ecuaciones canónicas de la recta que pasa por el punto dado* B *de coordenadas* $\beta_1, \ldots, \beta_n$ *y que es perpendicular al plano dado* (11).

Resolviendo conjuntamente las ecuaciones (11) y (14), encontramos las coordenadas

$$\xi_i = \frac{\alpha - (\alpha_1\beta_1 + \ldots + \alpha_n\beta_n)}{\alpha_1^2 + \ldots + \alpha_n^2}\alpha_i + \beta_i \quad (i = 1, \ldots, n)$$

de la proyección de B sobre el hiperplano (11).

32.4 Angulo entre un plano y una recta. Consideremos un plano arbitrario $\mathfrak{M}$ k-dimensional perteneciente a un espacio unitario $\mathfrak{U}$ de n dimensiones y una recta cualquiera $\mathfrak{X}$ de este espacio que intercepta el plano $\mathfrak{M}$ en el punto A. La proyección de la recta $\mathfrak{X}$ sobre el plano $\mathfrak{M}$ es una recta $\mathfrak{X}_{\mathfrak{M}}$ que pasa por el punto A. El ángulo entre la recta $\mathfrak{X}$ y su proyección $\mathfrak{X}_{\mathfrak{M}}$ se llama *ángulo entre la recta* $\mathfrak{X}$ *y el plano* $\mathfrak{M}$ (fig. 11). Si la recta $\mathfrak{X}$ se halla sobre el plano $\mathfrak{M}$, se tiene $\mathfrak{X} = \mathfrak{X}_{\mathfrak{M}}$ y el ángulo entre $\mathfrak{X}$ y $\mathfrak{M}$ es igual a cero. Si $\mathfrak{X}$ no se halla sobre el plano $\mathfrak{M}$, existe el único plano $(k+1)$-dimensional $\mathfrak{N}$ que pasa por $\mathfrak{M}$ y $\mathfrak{X}$. El plano $\mathfrak{M}$ es un hiperplano en el espacio $\mathfrak{N}$ y por ello se puede trazar en $\mathfrak{N}$ sólo una perpendicular $\mathfrak{Y}$ a $\mathfrak{M}$ por el punto A. Sea B un punto cualquiera de $\mathfrak{X}$ diferente de A y sea C la proyección de B sobre $\mathfrak{M}$. La recta $B \vee C$ se halla en el espacio $\mathfrak{N}$ y es perpendicular a $\mathfrak{M}$. Todas las perpendiculares a un mismo hiperplano son paralelas (véase el p. 32.3) y por esto las rectas $\mathfrak{Y}$, $\mathfrak{X}$ y $A \vee C$ (véase la fig. 11) se hallan sobre un mismo plano bidimensional. De aquí se deduce que *el ángulo* φ *entre* $\mathfrak{X}$ *y* $\mathfrak{M}$ *y el ángulo* ψ *entre* $\mathfrak{X}$ *y* $\mathfrak{Y}$ *dan en suma* $\pi/2$.

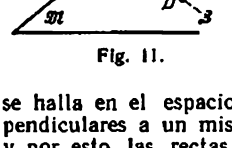


Fig. 11.

Resta considerar el caso en que el plano k-dimensional $\mathfrak{M}$ y la recta $\mathfrak{X}$ no se intersecan. Tomemos sobre el plano $\mathfrak{M}$ un punto cualquiera A. En el espacio $\mathfrak{U}$ existe, según el p. 30.3, una recta única $\mathfrak{X}'$ que pasa por el punto A y que es paralela a la recta $\mathfrak{X}$. El ángulo entre la recta $\mathfrak{X}'$ y el plano $\mathfrak{M}$ se llama en este caso ángulo entre la recta $\mathfrak{X}$ y el plano $\mathfrak{M}$. Es fácil comprobar que el ángulo entre la recta $\mathfrak{X}$ y el plano $\mathfrak{M}$ definido de esta forma no depende de cómo se escoja el punto auxiliar A.

Supongamos que en un espacio $\mathfrak{U}$ se ha fijado una base de referencia ortonormal $(O, e_1, \ldots, e_n)$ y que en esta base de referencia están dados un hiperplano $\mathfrak{M}$ por medio de la ecuación

$$\alpha_1\xi_1 + \ldots + \alpha_n\xi_n = \alpha \qquad (1)$$

y una recta $\mathfrak{X}$ mediante sus esuaciones canónicas

$$\frac{\xi_1-\xi_1^0}{l}=\ldots=\frac{\xi_n-\xi_n^0}{l_n}. \tag{2}$$

Tomemos en el hiperplano $\mathfrak{M}$ un punto A y sean $\sigma_1, \ldots, \sigma_n$ sus coordenadas. Entonces las ecuaciones de la recta $\mathfrak{X}'$ que es paralela a $\mathfrak{X}$ y que pasa por A son

$$\frac{\xi_1-\sigma_1}{l_1}=\ldots=\frac{\xi_n-\sigma_n}{l_n}, \tag{3}$$

y las ecuaciones de la perpendicular $\mathfrak{Y}$ levantada en el punto A hacia el hiperplano $\mathfrak{M}$ serán

$$\frac{\xi_1-\sigma_1}{\alpha_1}=\ldots=\frac{\xi_n-\sigma_n}{\alpha_n}. \tag{4}$$

Por lo visto en el p. 32.2, para el ángulo ψ entre $\mathfrak{X}'$ y $\mathfrak{Y}$ se tiene

$$\cos\psi=\frac{|\alpha_1 l_1+\ldots+\alpha_n l_n|}{\sqrt{|\alpha_1|^2+\ldots+|\alpha_n|^2}\ \sqrt{|l_1|^2+\ldots+|l_n|^2}}$$

y, por consiguiente, el ángulo φ entre $\mathfrak{M}$ y $\mathfrak{X}$ satisface las relaciones

$$\operatorname{sen}\varphi=\frac{|\alpha_1 l_1+\ldots+\alpha_n l_n|}{\sqrt{|\alpha_1|^2+\ldots+|\alpha_n|^2}\ \sqrt{|l_1|^2+\ldots+|l_n|^2}} \qquad \left(0\leqslant\varphi\leqslant\frac{\pi}{2}\right). \tag{5}$$

Igual que en el espacio euclídeo corriente, el ángulo entre una recta y un plano en un espacio unitario posee la propiedad de extremalidad.

TEOREMA. *Supongamos que en un espacio unitario* $\mathfrak{U}_n$ *de n dimensiones una recta* $\mathfrak{X}$ *se corta con un plano k-dimensional* $\mathfrak{M}$ *en un punto A formando un ángulo* φ *diferente de 0 y de* $\pi/2$. *Entonces el ángulo entre* $\mathfrak{X}$ *y cualquier recta* $\mathfrak{Z}$, *que pasa en el plano* $\mathfrak{M}$ *por A y es diferente de la proyección de* $\mathfrak{X}$ *sobre* $\mathfrak{M}$, *es mayor que* φ.

Tomemos sobre $\mathfrak{X}$ un punto cualquiera B diferente de A y tracemos desde B hacia $\mathfrak{M}$ una perpendicular indicando su base por C. Indiquemos por D la base de la perpendicular trazada desde el punto B hacia la recta $\mathfrak{Z}$ (véase la fig. 11). Puesto que la recta $B \vee C$ es perpendicular al plano $\mathfrak{M}$, los ángulos BCD y BCA son rectos. El ángulo BDA es recto por construcción. De los triángulos rectángulos BCA, BDA y BCD obtenemos, respectivamente

$$\operatorname{sen}\varphi=\frac{BC}{BA},\ \operatorname{sen}(\mathfrak{X},\mathfrak{Z})=\frac{BD}{BA}\ \text{y}\ BD=\sqrt{BC^2+DC^2}>BC.$$

Luego, $\operatorname{sen}\varphi<\operatorname{sen}(\mathfrak{X},\mathfrak{Z})$ y $\varphi<\angle(\mathfrak{X},\mathfrak{Z})$ que es lo que se quería demostrar

Ejemplos y problemas

En los problemas que se ofrecen a continuación se supone que en cada uno de los casos se ha escogido y se ha fijado una base de referencia ortonormal del espacio puntual euclídeo y que los puntos de este espacio se dan por las coordenadas calculadas en esta base

1. Hállese la proyección ortogonal del segmento $[(0, 0, 0, 0), (4, -1, -3, 4)]$ sobre el plano $\mathfrak{M} = A_0 \vee A_1 \vee A_2 \vee A_3$, donde las filas coordenadas de los puntos A_0, A_1, A_2 y A_3 son, respectivamente, (0, 0, 0, 0,), (1, 1, 1, 1), (1, 2, 2, −1) y (1, 0, 0, 3).

2. Se llama distancia de un punto A a un plano $\mathfrak{M}$ el mínimo de las distancias entre el punto dado y los puntos de $\mathfrak{M}$. Demuéstrese que esta distancia es igual a la longitud del vector $\overline{APr_{\mathfrak{M}}A}$.

3. Hállense las longitudes de los lados y los ángulos interiores del triángulo cuyos vértices están dados por las filas coordenadas (2, 4, 2, 4, 2), (6, 4, 4, 4, 6) y (5, 7, 5, 7, 2).

4. Hállese el ángulo entre la recta $A_0 \vee B$ y el plano $A_0 \vee A_1 \vee A_2$, donde las filas coordenadas de los puntos A_0, A_1, A_2, y B son iguales, respectivamente, a (0, 0, 0, 0), (3, 4, −4, −1), (0, 1, −1, 2) y (2, 2, 1, 1).

INDICE ALFABÉTICO

Adherencia 336, 380
Adjunto de un elemento 45
— de un menor 58
Algoritmo de Gauss 55
— de Lagrange 264
Angulo entre rayos 390
— — rectas 388
— — una recta y un plano 396
Antiautomorfismo 84
Antiendomorfismo 84
Antiisomorfismo 84
Aplicación 114, 121
— antisimétrica 233, 290, 299
— biyectiva 117
— conjugada 218, 288
— de Cayley 245, 299
— hermitiana 231
— idempotente 248
— idéntica 116, 122
— inducida 141
— inversa 116
— isométrica 293
— lineal 121, 240
— normal 220
— nula 122
— ortogonal 249, 306
— proyectiva 247
Aplicación regular 135
— seudounitaria 323
— simétrica 231, 235, 290, 208
— simplicial 313
— singular 135
— unitaria 225
Automorfismo 136, 293, 332
Base 75, 85, 110, 207, 338, 352
— de referencia 352, 388
Característica de Segre 182
— de una función bilineal 324
— Weyr 182
Célula de Jordan 149
— generalizada de Jordan 180
Circulador 48
Codimensión 340
Combinación lineal 71
Complemento ortogonal 213
— — a un plano 392
Congruencia de formas 257, 273
— de matrices 259, 273
Conjunto convexo 380, 382
Criterio de Jacobi 270
Cuaternio 29
Cuerpo ordenado 374
Defecto de una aplicación 133
— de una matriz 138
Dependencia lineal 73, 74, 335, 336
Descomposición espectral 249
— polar 243
Desigualdad de Bessel 209
— de Cauchy — Buniakovski 204
Desplazamiento 327
Determinante 33, 34, 35
— de Gram 215
— de una aplicación 143
— de Vandermonde 58
— volumen 355
Dimensión de un conjunto convexo 382
— de un espacio afín 338
— de un espacio lineal 77
Distancia entre puntos 386
— — vectores 206
Divisor elemental 166
— de un menor 158
Dominio de valores 133
Equivalencia de formas 257, 271
— de matrices 171, 229, 272
— de λ-matrices 154
Espacio afín sobre un cuerpo 328, 351
— — — un espacio vectorial 326
— bilineal métrico 282
— conjugado (o dual) 216
— euclídeo 201, 202, 286, 386
— de dimensión finita 76
— de filas 71
Espacio de funciones 203
— lineal (o vectorial) 69, 71
— métrico 386
— seudoeuclídeo 285
— seudounitario 286, 315
— simplicial 285, 309
— unitario 202, 386
Espectro de una aplicación lineal 251
Factores invariantes 161, 272
Fila 71
— coordenada 86, 353, 354
Forma 256
— bilineal 256, 258, 259, 264
— — hermitiana 258, 259, 263
— canónica de una λ-matriz 165
— cuadrática 284, 260
— de Jordan 150, 180
— de orden s 358
— lineal 254
— normal natural 177
— polilineal 256
Fórmulas de Cramer 55
Función 72
— antilineal 225
— bilineal 278, 280, 281
— cuadrática 280, 281
— lineal 215
— de matriz 185
Funcional lineal 346
— de orden s 358
— polilineal 358
Hiperplano 104, 341, 362, 385
Hipersuperficie algebraica 356
Homomorfismo 332, 351
Identidad de Lagrange 51
Igualdad de Parseval 210
Independencia lineal 73, 335, 336
Isomorfismo de aplicaciones lineales 135, 228, 294
— de espacios lineales 81, 135
— de espacios bilineales métricos 284
— de espacios unitarios 211
Ley de dualidad 351

— de inercia 261, 268
Longitud de un vector 204
Matriz 13
— adjunta 61
— antisimétrica 21
— asociada 177
— canónica diagonal 155
— característica 60
— celular diagonal 27
— celular triangular 65
— conjugada 24
— de Gram 215, 282
Matriz de Jordan 150, 174
— del cambio 86, 370
— descompuesta 65
— de una aplicación inducida 141
— de una aplicación lineal 124
— de una forma 257
— de una función bilineal 279
— de una transformación 255
— de un automorfismo 372
— de un sistema 33, 34, 54, 97
— escalar 171
— finita en fila 33
hermitiana 24, 234
— idempotente 32
— inversa 22, 88
— involutiva 32
— nilpotente 67
— nula 14
— ortogonal 23, 52
— polinomial (λ-matriz) 153
— regular 52
— semidescompuesta 37, 65, 140
— semisimple 183
— simétrica 21
— simplicial 313
— singular 52
— transpuesta 21, 43
— triangular 66
Matriz unitaria 24, 53, 226
Menor 45, 58, 67
Métrica bilineal 282
Módulo 68
— unitario 69, 84
Multiplicidad de un valor propio 146
Núcleo de una aplicación 133
Número característico 61
Orden dual 375
— natural sobre una recta 375
Ortogonalidad de conjuntos de vectores 212
— de subespacios 212
— de vectores 206, 283
Plano 104, 336
— complementario 340
— coordenado r-dimensional 362
— vacío 338
Poliedro convexo 385
Polinomio característico de un aplicación 143
— — de una matriz 60
— matricial 20, 170
— mínimo de una aplicación 144
— — de una matriz 63
Proceso de Gram-Schmidt 210
Producto de aplicaciones 115
— de matrices 16
— escalar 202, 203
Proyección de un vector sobre un subespacio 214, 247
— ortogonal 393
Quebrada 387
Rango de una aplicación 133
— de una forma 257
— de una matriz 89
— de un sistema homogéneo 110
Radio vector de un punto 327
Rayo 376
Recta 104, 338
Regla del triángulo 330
Representación de una funcional 353
Segmento 376
Semejanza de aplicaciones 136, 228
— de matrices 58, 172
Semiespacio 378
Semiplano 379
Serie de potencias de una matriz 193
Signatura de una forma 260
— de un espacio 285, 286
Símplice 385
Sistema convexamente irreducible de puntos 385
— de coordenadas 85
— de ecuaciones 356
— de ecuaciones lineales 53, 54, 98, 109
— de generadores 75
— fundamental de soluciones 110
— linealmente dependiente de vectores 73
— normal de coordenadas 301, 309, 315, 316
— ortogonal de vectores 207
— ortonormal de vectores 208
Sistema simplicial de coordenadas 309
Solución de un sistema de ecuaciones lineales 53, 97
— nula (o trivial) 110
Subespacio 102
— complementario 340
— invariante 139
— isótropo 283
— lineal 102
— nulo 102
— propio (o no trivial) 102
— radical 151
— tangente 344
— trivial 102
Suma de aplicaciones 130
— de una serie de potencias de una matriz 193
— de subespacios 104, 107, 212
— directa de matrices 27
Sustitución 118, 119
Teorema de Hamilton — Cayley 62
— de Kronecker — Capelli 98
— de Schur 239
Transformación de coordenadas 87, 125
— elemental de λ-matrices 154
Traslación 327
Trasladó 327
Traza de una aplicación 143
— de una matriz 61
Valor propio de una aplicación 144
— — de una matriz 61
Variedad algebraica 357
— de álgebras 351
— lineal 336
— radical 348
Vector 69
— isótropo 283
— libre 326
— propio 144
— radical 151
— unitario (normalizado) 204